PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/12, C07K 14/46, 14/705, 16/18, C12N 15/11, A01K 67/027, C12Q 1/68, G01N 33/68, 33/566, A61K 38/17, C12N 15/62

A1

(11) International Publication Number: **WO 98/12326**

(43) International Publication Date: 26 March 1998 (26.03.98)

(21) International Application Number: PCT/US97/16741

(22) International Filing Date: 19 September 1997 (19.09.97)

(30) Priority Data:
60/026,155 20 September 1996 (20.09.96) US

(71) Applicant *(for all designated States except US)*: PRESIDENT AND FELLOWS OF HARVARD COLLEGE [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: MCMAHON, Andrew, P. [GB/US]; 128 Kendall Road, Lexington, MA 02173 (US). CHUANG, Pao, Tien [-/US]; 23 Roberts Road, Cambridge, MA 02138 (US).

(74) Agents: VINCENT, Matthew, P. et al.; Foley, Hoag & Eliot LLP, One Post Office Square, Boston, MA 02109 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: HEDGEHOG INTERACTING PROTEINS AND USES RELATED THERETO

HIP-1 (Hedgehog-Interacting Protein-1)

- Signal peptide
- EGF repeat
- Potential N-linked glycosylation site
- Transmembrane domain

(57) Abstract

The present invention concerns the discovery of a new family of *hedgehog* binding proteins, referred to herein as "*hedgehog* interacting proteins" or "*HIPs*", which are demonstrated to bind to *hedgehog* polypeptides with high affinity. As described herein, the vertebrate *HIP* proteins exhibit spatially and temporally restricted expression domains indicative of important roles in *hedgehog*-mediated induction.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	R	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

Kandel, Schwartz and Jessell, Elsevier Science Publishing Company: NY, 1991; and *Developmental Biology (3rd)*, ed. S.F. Gilbert, Sinauer Associates: Sunderland MA, 1991). Inductive interactions that define the fate of cells within the neural tube establish the initial pattern of the embryonic vertebrate nervous system. In the spinal cord, the identify of cell types is controlled, in part, by signals from two midline cell groups, the notochord and floor plate, that induce neural plate cells to differentiate into floor plate, motor neurons, and other ventral neuronal types (van Straaten et al. (1988) *Anat. Embryol.* 177:317-324; Placzek et al. (1993) *Development* 117:205-218; Yamada et al. (1991) *Cell* 64:035-647; and Hatta et al. (1991) *Nature* 350:339-341). In addition, signals from the floor plate are responsible for the orientation and direction of commissural neuron outgrowth (Placzek, M. et al., (1990) *Development* 110: 19-30). Besides patterning the neural tube, the notochord and floorplate are also responsible for producing signals which control the patterning of the somites by inhibiting differentiation of dorsal somite derivatives in the ventral regions (Brand-Saberi, B. et al., (1993) *Anat. Embryol.* 188: 239-245; Porquie, O. et al., (1993) *Proc. Natl. Acad. Sci. USA* 90: 5242-5246).

Another important signaling center exists in the posterior mesenchyme of developing limb buds, called the Zone of Polarizing Activity, or "ZPA". When tissue from the posterior region of the limb bud is grafted to the anterior border of a second limb bud, the resultant limb will develop with additional digits in a mirror-image sequence along the anteroposterior axis (Saunders and Gasseling, (1968) *Epithelial-Mesenchymal Interaction*, pp. 78-97). This finding has led to the model that the ZPA is responsible for normal anteroposterior patterning in the limb. The ZPA has been hypothesized to function by releasing a signal, termed a "morphogen", which forms a gradient across the early embryonic bud. According to this model, the fate of cells at different distances from the ZPA is determined by the local concentration of the morphogen, with specific thresholds of the morphogen inducing successive structures (Wolpert, (1969) *Theor. Biol.* 25:1-47). This is supported by the finding that the extent of digit duplication is proportional to the number of implanted ZPA cells (Tickle, (1981) *Nature* 254:199-202).

Although the existence of inductive signals in the ZPA has been known for years, the molecular identities of these signals are only now beginning to be elucidated. An important step forward has been the discovery that the secreted protein *Sonic hedgehog* (*Shh*) is produced in several tissues with organizing properties, including notochord, floor plate and ZPA (Echelard et al. (1993), *Cell* 75: 1417-1430; Bitgood, M.J. and A.P. McMahon (1995) *Dev. Biol.* 172:126-38). Misexpressing *Shh* mimics the inductive effects on ectopic notochord in the neural tube and somites (Echelard et al. (1993) *supra*) and also mimics ZPA function in the limb bud (Riddle et al. (1993) *Cell* 75:1401-16; Chang et al. (1994) *Development* 120: 3339-53).

Hedgehog Interacting Proteins and Uses Related Thereto

Funding

Work described herein was supported by funding from the National Institutes of Health. The United States Government has certain rights in the invention.

Background of the Invention

Pattern formation is the activity by which embryonic cells form ordered spatial arrangements of differentiated tissues. The physical complexity of higher organisms arises during embryogenesis through the interplay of cell-intrinsic lineage and cell-extrinsic signaling. Inductive interactions are essential to embryonic patterning in vertebrate development from the earliest establishment of the body plan, to the patterning of the organ systems, to the generation of diverse cell types during tissue differentiation (Davidson, E., (1990) *Development* 108: 365-389; Gurdon, J. B., (1992) *Cell* 68: 185-199; Jessell, T. M. et al., (1992) *Cell* 68: 257-270). The effects of developmental cell interactions are varied. Typically, responding cells are diverted from one route of cell differentiation to another by inducing cells that differ from both the uninduced and induced states of the responding cells (inductions). Sometimes cells induce their neighbors to differentiate like themselves (homoiogenetic induction); in other cases a cell inhibits its neighbors from differentiating like itself. Cell interactions in early development may be sequential, such that an initial induction between two cell types leads to a progressive amplification of diversity. Moreover, inductive interactions occur not only in embryos, but in adult cells as well, and can act to establish and maintain morphogenetic patterns as well as induce differentiation (J.B. Gurdon (1992) *Cell* 68:185-199).

The origin of the nervous system in all vertebrates can be traced to the end of gastrulation. At this time, the ectoderm in the dorsal side of the embryo changes its fate from epidermal to neural. The newly formed neuroectoderm thickens to form a flattened structure called the neural plate which is characterized, in some vertebrates, by a central groove (neural groove) and thickened lateral edges (neural folds). At its early stages of differentiation, the neural plate already exhibits signs of regional differentiation along its anterior posterior (A-P) and mediolateral axis (M-L). The neural folds eventually fuse at the dorsal midline to form the neural tube which will differentiate into brain at its anterior end and spinal cord at its posterior end. Closur of the neural tube creates dorsal/ventral differences by virtue of previous mediolateral differentiation. Thus, at the end of neurulation, the neural tube has a clear anterior-posterior (A-P), dorsal ventral (D-V) and mediolateral (M-L) polarities (see, for example, *Principles in Neural Science (3rd)*, eds.

The vertebrate family of *hedgehog* genes includes at least four members, e.g., paralogs of the single drosophila *hedgehog* gene. Exemplary hedgehog genes and proteins are described in PCT publications WO 95/18856 and WO 96/17924. Three of these members, herein referred to as Desert *hedgehog* (*Dhh*), Sonic *hedgehog* (*Shh*) and Indian *hedgehog* (*Ihh*), apparently exist in all vertebrates, including fish, birds, and mammals. A fourth member, herein referred to as tiggie-winkle *hedgehog* (*Thh*), appears specific to fish. Desert hedgehog (*Dhh*) is expressed principally in the testes, both in mouse embryonic development and in the adult rodent and human; Indian hedgehog (*Ihh*) is involved in bone development during embryogenesis and in bone formation in the adult; and, *Shh*, which as described above, is primarily involved in morphogenic and neuroinductive activities. Given the critical inductive roles of hedgehog polypeptides in the development and maintenance of vertebrate organs, the identification of hedghog interacting proteins is of paramount significance in both clinical and research contexts.

Summary of the Invention

The present invention relates to the discovery of a new class of *hedgehog*-binding protein, referred to herein as *HIP* (for *hedgehog* interacting protein). The *HIP* polypeptides of the present invention include polypeptides which bind the products of the *hedgehog* gene family. *Hedgehog* family members are known for their broad involvement in the formation and maintenance of ordered spatial arrangements of differentiated tissues in vertebrates, both adult and embryonic, and can be used to generate and/or maintain an array of different vertebrate tissue both *in vitro* and *in vivo*.

In general, the invention features isolated *HIP* polypeptides, preferably substantially pure preparations of the subject *HIP* polypeptides. The invention also provides recombinantly produced *HIP* polypeptides. In preferred embodiments the polypeptide has a biological activity including the ability to bind a *hedgehog* protein with high affinity, e.g., with a nanomolar or smaller dissociation constant (K_D). *HIP* polypeptides which specifically antagonize such activities, such as may be provided by truncation mutants, are also specifically contemplated.

In one embodiment, the polypeptide is identical with or homologous to a *HIP* polypeptide represented in SEQ ID No: 5, SEQ ID No: 6, SEQ ID No: 7 and SEQ ID No: 8, or the core polypeptide sequence thereof (e.g., corresponding to residues 16-678 of SEQ ID. 5 or 6). Related members of the *HIP* family are also contemplated, for instance, a *HIP* polypeptide preferably has an amino acid sequence at least 65%, 67%, 69%, 70%, 75% or 80% homologous to a polypeptide represented by SEQ ID No: 5, SEQ ID No: 6, SEQ ID No: 7 and SEQ ID No: 8 though polypeptides with higher sequence homologies of, for example, 82%, 85%, 90% and 95% or are also contemplated. In a preferred embodiment,

the *HIP* polypeptide is encoded by a nucleic acid which hybridizes under stringent conditions with a nucleic acid sequence represented in any one or more of SEQ ID Nos: 1-4 and 9-14. Homologs of the subject *HIP* proteins also include versions of the protein which are resistant to post-translation modification. as for example, due to mutations which alter modification sites (such as tyrosine, threonine, serine or aspargine residues), or which prevent glycosylation of the protein, or which prevent interaction of the protein with a *HIP* ligand, e.g. a *hedgehog* polypeptide.

The *HIP* polypeptide can comprise a full length protein, such as represented in SEQ ID No: 5, SEQ ID No: 6 or SEQ ID No: 7, or it may include the core polypeptide sequence thereof (e.g., corresponding to residues 16-678 of SEQ ID. 5 or 6), or it can include a fragment corresponding to one or more particular motifs/domains, or to arbitrary sizes, e.g., at least 5, 10, 25, 50, 100, 150 or 200 amino acids in length. In preferred embodiments, the *HIP* polypeptide includes a sufficient portion of the excellular ligand binding domain to be able to specifically bind to a *hedgehog* ligand, preferably with a K_D of 9μM or less and even more preferably of 9nM or less. Truncated forms of the protein include, but are not limited to, soluble ligand binding domain fragments.

In certain preferred embodiments, the invention features a purified or recombinant *HIP* polypeptide having a core polypeptide molecular weight of about 78.4kd. In other embodiments, the peptide core of a mature *HIP* protein preferably has a molecular weight in the range of 38.6 to 76.8kD. It will be understood that certain post-translational modifications, e.g., glycosylation, prenylation, myristylation and the like, can increase the apparent molecular weight of the *HIP* protein relative to the unmodified polypeptide chain.

The subject proteins can also be provided as chimeric molecules, such as in the form of fusion proteins. For instance, the *HIP* protein can be provided as a recombinant fusion protein which includes a second polypeptide portion, e.g., a second polypeptide having an amino acid sequence unrelated (heterologous) to the *HIP* polypeptide, e.g. the second polypeptide portion is glutathione-S-transferase, e.g. the second polypeptide portion is an enzymatic activity such as alkaline phosphatase, e.g. the second polypeptide portion is an epitope tag.

In yet another embodiment, the invention features nucleic acids encoding *HIP* polypeptides, which have the ability to modulate, e.g., either mimic or antagonize, at least a portion of the activity of a wild-type *HIP* polypeptide. Exemplary *HIP*-encoding nucleic acid sequences are represented by SEQ ID No: 1, SEQ ID No: 2, SEQ ID No: 3 or SEQ ID No: 4.

In another embodiment, the nucleic acids of the present invention include coding sequences which hybridize under stringent conditions with all or a portion of the coding sequences designated in one or more of SEQ ID Nos: 1-4. The coding sequences of the

nucleic acids can comprise sequences which are identical to coding sequences represented in SEQ ID Nos: 1, 2, 3, 4, 9, 10, 11, 12, 13 or 14, or it can merely be homologous to those sequences. In preferred embodiments, the nucleic acids encode polypeptides which specifically modulate, by acting as either agonists or antagonists, one or more of the bioactivities of wild-type *HIP* polypeptides.

Furthermore, in certain preferred embodiments, the subject *HIP* nucleic acids will include a transcriptional regulatory sequence, e.g. at least one of a transcriptional promoter or transcriptional enhancer sequence, which regulatory sequence is operably linked to the *HIP* gene sequences. Such regulatory sequences can be used in to render the *HIP* gene sequences suitable for use as an expression vector. The transcriptional regulatory sequence can be from a HIP gene, or from a heterologous gene.

This invention also contemplates the cells transfected with said expression vector whether prokaryotic or eukaryotic and a method for producing *HIP* proteins by employing said expression vectors.

In still other embodiments, the subject invention provides a gene activation construct, wherein the gene activation construct is deigned to recombine with a genomic *HIP* gene in a cell to provide, e.g., by heterologous recombination, a heterologous transcriptional regulatory sequence operatively linked to a coding sequence of a genomic *HIP* gene. Cells having genomic *HIP* genes modified by gene activation constructs are also specifically contemplated.

In yet another embodiment, the present invention provides nucleic acids which hybridize under stringent conditions to nucleic acid probes corresponding to at least 12 consecutive nucleotides of either sense or antisense sequences of SEQ ID No: 1, SEQ ID No: 2, SEQ ID No: 3 and SEQ ID No: 4; though preferably to at least 25 consecutive nucleotides; and more preferably to at least 40, 50 or 75 consecutive nucleotides of either sense or antisense sequence of SEQ ID No: 1, SEQ ID No: 2, SEQ ID No: 3 and SEQ ID No: 4.

Yet another aspect of the present invention concerns an immunogen comprising a *HIP* polypeptide in an immunogenic preparation, the immunogen being capable of eliciting an immune response specific for a *HIP* polypeptide; e.g. a humoral response, e.g. an antibody response; e.g. a cellular response. In preferred embodiments, the immunogen comprising an antigenic determinant, e.g. a unique determinant, from a protein represented by one of SEQ ID No: 5, SEQ ID No: 6, SEQ ID No: 7 and/or SEQ ID No: 8.

A still further aspect of the present invention features antibodies and antibody preparations specifically reactive with an epitope of the *HIP* immunogen.

The invention also features transgenic non-human animals, e.g. mice, rats, rabbits, chickens, frogs or pigs, having a transgene, e.g., animals which include (and preferably express) a heterologous form of a *HIP* gene described herein, or which misexpress an endogenous *HIP* gene, e.g., an animal in which expression of one or more of the subject *HIP* proteins is disrupted. Such a transgenic animal can serve as an animal model for studying cellular and tissue disorders comprising mutated or mis-expressed *HIP* alleles or for use in drug screening.

The invention also provides a probe/primer comprising a substantially purified oligonucleotide, wherein the oligonucleotide comprises a region of nucleotide sequence which hybridizes under stringent conditions to at least 12 consecutive nucleotides of sense or antisense sequences of any one or more of SEQ ID Nos: 1-4 and 9-14, or naturally occurring mutants thereof. In preferred embodiments, the probe/primer further includes a label group attached thereto and able to be detected. The label group can be selected, e.g., from a group consisting of radioisotopes, fluorescent compounds, enzymes, and enzyme co-factors. Probes of the invention can be used as a part of a diagnostic test kit for identifying dysfunctions associated with mis-expression of a *HIP* protein, such as for detecting in a sample of cells isolated from a patient, a level of a nucleic acid encoding a *HIP* protein; e.g. measuring a *HIP* mRNA level in a cell, or determining whether a genomic *HIP* gene has been mutated or deleted. These so-called "probes/primers" of the invention can also be used as a part of "antisense" therapy which refers to administration or *in situ* generation of oligonucleotide probes or their derivatives which specifically hybridize (e.g. bind) under cellular conditions, with the cellular mRNA and/or genomic DNA encoding one or more of the subject *HIP* proteins so as to inhibit expression of that protein, e.g. by inhibiting transcription and/or translation. Preferably, the oligonucleotide is at least 12 nucleotides in length, though primers of 25, 40, 50, or 75 nucleotides in length are also contemplated.

In yet another aspect, the invention provides an assay for screening test compounds for inhibitors, or alternatively, potentiators, of an interaction between a *hedgehog* protein and a *HIP* polypeptide receptor. An exemplary method includes the steps of (a) forming a reaction mixture including: (i) a *hedgehog* polypeptide, (ii) a *HIP* polypeptide, and (iii) a test compound; and (b) detecting interaction of the *hedgehog* and *HIP* polypeptides. A statistically significant change (potentiation or inhibition) in the interaction of the *hedgehog* and *HIP* polypeptides in the presence of the test compound, relative to the interaction in the absence of the test compound, indicates a potential agonist (mimetic or potentiator) or antagonist (inhibitor) of *hedgehog* bioactivity for the test compound. The reaction mixture can be a cell-free protein preparation, e.g., a reconsistuted protein mixture or a cell lysate, or it can be a recombinant cell including a heterologous nucleic acid recombinantly expressing the *HIP* polypeptide.

In preferred embodiments, the step of detecting interaction of the *hedgehog* and *HIP* polypeptides is a competitive binding assay. In other preferred embodiments, the step of detecting interaction of the *hedgehog* and *HIP* polypeptides involves detecting, in a cell-based assay, change(s) in the level of an intracellular second messenger responsive to signaling mediated by the *HIP* polypeptide. In still another preferred embodiment, the step of detecting interaction of the *hedgehog* and *HIP* polypeptides comprises detecting, in a cell-based assay, change(s) in the level of expression of a gene controlled by a transcriptional regulatory sequence responsive to signaling by the *HIP* polypeptide.

In preferred embodiments, the steps of the assay are repeated for a variegated library of at least 100 different test compounds, more preferably at least 10^3, 10^4 or 10^5 different test compounds. The test compound can be, e.g., a peptide, a nucleic acid, a carbohydrate, a small organic molecule, or natural product extract (or fraction thereof).

The present invention further contemplates the pharmaceutical formulation of one or more agents identified in such drug screening assays.

In other embodiments, the present invention provides a molecule, preferably a small organic molecule, which binds to *HIP* and either mimics or antagonizes *hedgehog*-induced signaling in cells expressing *HIP*.

Yet another aspect of the present invention concerns a method for modulating one or more of growth, differentiation, or survival of a cell by modulating *HIP* bioactivity, e.g., by potentiating or disrupting certain protein-protein interactions. In general, whether carried out *in vivo*, *in vitro*, or *in situ*, the method comprises treating the cell with an effective amount of a *HIP* therapeutic so as to alter, relative to the cell in the absence of treatment, at least one of (i) rate of growth, (ii) differentiation, or (iii) survival of the cell. Accordingly, the method can be carried out with *HIP* therapeutics such as peptide and peptidomimetics or other molecules identified in the above-referenced drug screens which agonize or antagonize the effects of signaling from a *HIP* protein or ligand binding of a *HIP* protein, e.g., a *hedgehog* protein. Other *HIP* therapeutics include antisense constructs for inhibiting expression of *HIP* proteins, dominant negative mutants of HIP proteins which competitively inhibit ligand interactions upstream and signal transduction downstream of the wild-type HIP protein, and gene therapy constructs including gene activation constructs.

In one embodiment, the subject method of modulating *HIP* bioactivity can be used in the treatment of testicular cells, so as to modulate spermatogenesis. In another embodiment, the subject method is used to modulate osteogenesis, comprising the treatment of osteogenic cells with an agent that modulates *HIP* boactivity. Likewise, where the treated cell is a chondrogenic cell, the present method is used to modulate chondrogenesis. In still, another embodiment, the subject method can be used to modulate the differentiation of a neuronal cell, to maintain a neuronal cell in a differentiated state, and/or to enhance the

survival of a neuronal cell, e.g., to prevent apoptosis or other forms of cell death. For instance the present method can be used to affect the differentiation of neuronal cells such as motor neurons, cholinergic neurons, dopaminergic neurons, serotonergic neurons, and peptidergic neurons.

Another aspect of the present invention provides a method of determining if a subject, e.g. an animal patient, is at risk for a disorder characterized by unwanted cell proliferation or aberrant control of differentiation or apoptosis. The method includes detecting, in a tissue of the subject, the presence or absence of a genetic lesion characterized by at least one of (i) a mutation of a gene encoding a *HIP* protein; or (ii) the mis-expression of a *HIP* gene. In preferred embodiments, detecting the genetic lesion includes ascertaining the existence of at least one of: a deletion of one or more nucleotides from a *HIP* gene; an addition of one or more nucleotides to the gene, a substitution of one or more nucleotides of the gene, a gross chromosomal rearrangement of the gene; an alteration in the level of a messenger RNA transcript of the gene; the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; a non-wild type level of the protein; and/or an aberrant level of soluble *HIP* protein.

For example, detecting the genetic lesion can include (i) providing a probe/primer including an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence of a *HIP* gene or naturally occurring mutants thereof, or 5' or 3' flanking sequences naturally associated with the *HIP* gene; (ii) exposing the probe/primer to nucleic acid of the tissue; and (iii) detecting, by hybridization of the probe/primer to the nucleic acid, the presence or absence of the genetic lesion; e.g. wherein detecting the lesion comprises utilizing the probe/primer to determine the nucleotide sequence of the *HIP* gene and, optionally, of the flanking nucleic acid sequences. For instance, the probe/primer can be employed in a polymerase chain reaction (PCR) or in a ligation chain reaction (LCR). In alternate embodiments, the level of a *HIP* protein is detected in an immunoassay using an antibody which is specifically immunoreactive with the *HIP* protein.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, Molecular Cloning A Laboratory Manual, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989); DNA Cloning, Volumes I and II (D. N. Glover ed., 1985); Oligonucleotide Synthesis (M. J. Gait ed., 1984); Mullis et al. U.S. Patent No: 4,683,195; Nucleic Acid Hybridization (B. D. Hames & S. J. Higgins eds. 1984); Transcription And Translation (B. D. Hames & S. J. Higgins eds. 1984); Culture Of Animal Cells (R. I. Freshney, Alan R. Liss, Inc., 1987); Immobilized Cells And Enzymes (IRL

Press. 1986); B. Perbal, A Practical Guide To Molecular Cloning (1984); the treatise, Methods In Enzymology (Academic Press, Inc., N.Y.); Gene Transfer Vectors For Mammalian Cells (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); Methods In Enzymology, Vols. 154 and 155 (Wu et al. eds.), Immunochemical Methods In Cell And Molecular Biology (Mayer and Walker, eds., Academic Press, London, 1987); Handbook Of Experimental Immunology, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

Brief Description of the Drawings

Figure 1A is an alignment of the *HIP* protein sequences for the mouse, human, chicken and zebrafish homologs. The up-arrow indicates the C-terminal hyrdophobic anchor.

Figure 1B is an alignment of the coding sequences for *HIP* cDNAs isolated from mouse, human, chicken and zebrafish.

Figure 2 is a schematic representation of the HIP protein.

Figure 3 shows two scatchard plots of the binding of a Shh-AP fusion protein (Ap=alkaline phosphatase) with HIP and PTC proteins.

Figure 4 is a human multiple tissue Northern blot for HIP transcripts.

Figure 5 is a mouse multiple tissue Northern blot for HIP transcripts.

Figure 6 illustrates that truncated forms of the HIP protein, in this instance lacking the C-terminal 22 amino acids, are secreted into the cell supernatant, whereas the full length HIP protein is retained in the cell fraction, e.g., remains membrane bound. Moreover, in the presence of Shh, anti-Shh can immunoprecipitate a complex including the secreted form of HIP protein.

Detailed Description of the Invention

Of particular importance in the development and maintenance of tissue in vertebrate animals is a type of extracellular communication called induction, which occurs between neighboring cell layers and tissues. In inductive interactions, chemical signals secreted by one cell population influence the developmental fate of a second cell population. Typically, cells responding to the inductive signals are diverted from one cell fate to another, neither of which is the same as the fate of the signaling cells.

Inductive signals are key regulatory proteins that function in vertebrate pattern formation, and are present in important signaling centers known to operate embryonically, for example, to define the organization of the vertebrate embryo. For example, these signaling structures include the notochord, a transient structure which initiates the formation of the nervous system and helps to define the different types of neurons within it. The notochord also regulates mesodermal patterning along the body axis. Another distinct group of cells having apparent signaling activity is the floorplate of the neural tube (the precursor of the spinal cord and brain) which also signals the differentiation of different nerve cell types. It is also generally believed that the region of mesoderm at the bottom of the buds which form the limbs (called the Zone of Polarizing Activity or ZPA) operates as a signaling center by secreting a morphogen which ultimately produces the correct patterning of the developing limbs.

The regulation of *hedgehog* protein signaling is an important mechanism for developmental control. The present invention concerns the discovery of a new family of *hedgehog* binding proteins, refered to herein as "*hedgehog* interacting proteins" or "*HIPs*", which are demonstrated to bind to *hedgehog* polypeptides with high affinity. The mouse HIP clone was first identified by expression cloning techniques by its ability to bind to *hedgehog* protein. Subsequently, a variety of other vertebrate homologs have been cloned using probes and primers based on the mouse clone, again by standard techniques. As described herein, the vertebrate *HIP* proteins exhibit spatially and temporally restricted expression domains indicative of important roles in *hedgehog*-mediated induction.

The sequence of exemplary *HIP* genes cloned from various vertebrates (*c.f.*, Table 1 below) indicates it encodes a secreted protein that may be anchored at the cell membrane. Comparison of *HIP* sequences from mouse, human, chick and zebrafish (see Figure 1) suggests a conserved signal peptide sequence, a conserved *hedgehog* binding domain, and a potential transmembrane domain. Moreover, analysis of the protein sequences suggests 2 EGF-like domains in the C-terminal portion of the protein (see Figure 2). Other than those domains, the *HIP* coding sequences do not show close sequence homology to any previously identified genes, suggesting that these genes comprise a novel gene family.

The *HIP* proteins, through their ability to bind to *hedgehog* proteins, are apparently capable of modulating *hedgehog* signaling. The *HIP* proteins may function as a *hedgehog* receptor (or subunit thereof), or may act to sequester *hedgehog* proteins at the cell surface and thus control the effective concentration of *hedgehog* polypeptide available to other *hedgehog* receptors such as *patched*. The *HIP* proteins may mediate formation of a *hedgehog* gradient by forming complexes with soluble *hedgehog* proteins and affecting the ability of those proteins to interact with cell-surface receptors. Thus, the *HIP* polypeptides of the present invention may affect a number of *hedgehog*-mediated biological activities

including: an ability to modulate proliferation, survival and/or differentiation of mesodermally-derived tissue, such as tissue derived from dorsal mesoderm, cartilage and tissue involved in spermatogenesis; the ability to modulate proliferation, survival and/or differentiation of ectodermally-derived tissue, such as tissue derived from the epidermis, neural tube, neural crest, or head mesenchyme; the ability to modulate proliferation, survival and/or differentiation of endodermally-derived tissue, such as tissue derived from the primitive gut.

A mouse *HIP* cDNA was identified in a screen for potential *hedgehog*-binding proteins using a mouse limb bud cDNA library cloned into a plasmid which allowed expression in cells, and detecting the amount of labeled *Shh* protein that bound specifically to the expressed proteins. A single positive cone was identified in 70,000 screened. Ligand-receptor binding studies indicate that the *HIP* polypeptide can bind various members of the *hedgehog* family with high affinity. For instance, the binding of the murine *HIP* polypeptide to each of *Shh* and *Dhh* occurred with a dissociation constant (k_d) of approximately 1nM. For example, see Figure 3. This binding is comparable to the *hedgehog* binding affinity observed for patched (see Figure 3). This finding suggests that mouse *HIP* cDNA may encode a general *hedgehog* binding protein as opposed to a binding protein that selectively discriminates between *hedgehog* homologs. However, it is anticipated that other homologs of that protein may be able to distinguish, by binding affinity, between *Shh, Ihh* and *Dhh*.

In addition to the murine *HIP* clone, we have also obtained cDNA clones from other vertebrates, including human, avian and fish *HIP* genes, utilizing the mouse cDNA as a probe. According to the appended sequence listing, (see also Table 1) a murine *HIP* polypeptide is encoded by SEQ ID No:1; a human *HIP* polypeptide is encoded by SEQ ID No:2; a chicken *HIP* polypeptide is encoded by SEQ ID No:3; and a zebrafish *HIP* polypeptide is encoded by SEQ ID No:4.

Table 1

Guide to *HIP* sequences in Sequence Listing

	Nucleotide	Amino Acid
Mouse *HIP*	SEQ ID No. 1	SEQ ID No. 5
Human *HIP*	SEQ ID No. 2	SEQ ID No. 6
5' partial	SEQ ID No. 9	
internal	SEQ ID No. 10	
3' partial	SEQ ID No. 11	

Chicken *HIP*	SEQ ID No. 3	SEQ ID No. 7
5' partial	SEQ ID No. 12	
internal	SEQ ID No. 13	
3' partial	SEQ ID No. 14	
Zebrafish *HIP*	SEQ ID No. 4	SEQ ID No. 8

The overall sequence homology between the *HIP* proteins is shown in Table 2.

Table 2

Amino acid sequence identity between vertebrate *HIP* proteins.

	Mouse	Human	Chicken
Human	95%		
Chicken	82%	85%	
Fish	69%	69%	67%

By fluorescence *in situ* hybridization (FISH), a human *HIP* gene has been localized to chromosomal position 4Q31. As illustrated in Figures 4 and 5, Northern blot analysis suggests that a *HIP* gene is expressed in certain adult tissues, with higher levels indicated in heart, skeletal muscle and pancreas, at least in the tissue samples tested to date.

It is contemplated by the present invention that the cloned *HIP* genes set out in the appended sequence listing, in addition to representing a inter-species family of related genes, are also each part of an intra-species family. That is, it is anticipated that other paralogs of the human and mouse *HIP* proteins exist in those animals, and orthologs of each *HIP* gene are conserved amongst other animals. For instance, at low to medium stringency conditions, transcripts of about 4.4kb and 9 kb were observed by Northern analysis of mouse samples (see Figure 5), the latter representing a likely paralog and/or splice variant of the HIP cDNA set forth in SEQ ID No. 1.

In addition to the sequence variation between the various HIP homologs, the vertebrate HIP proteins are apparently present naturally in a number of different forms, including a pro-form. The pro-form includes an N-terminal signal peptide (approximately N-terminal residues 1-15) for directed secretion of at least the N terminal domain of the protein, while the full-length mature form lacks this signal sequence. Further processing of the mature form may also occur in some instances to yield biologically active fragments of the protein.

Likewise, as illustrated in Figure 6, the full-length *HIP* protein also includes a membrane anchor domain, e.g., a transmembrane domain, comprised of about the C-terminal 22 amino acids of the protein. *HIP* polypeptides lacking this sequence are shown to be fully secreted rather than membrane bound. Briefly, a myc-tagged fusion protein was created with the full length HIP sequence, myc-HIP-1, and a truncated form of HIP missing the C-terminal 22 amino acids, myc-HIP-1(Δ22). The myc-HIP-1 fusion protein was shown to run just slightly slower (high MW) than the full-length HIP protein when each was detected by anti-myc and anti-HIP antibodies, respectively. The anti-myc antibody was used to immunobot samples of cell pellets and cell supernatant produced by cells expressing either the myc-HIP-1 fusion protein or the myc-HIP-1(Δ22) fusion protein. For the cells expressing myc-HIP-1, e.g., which retains the putative membrane anchoring domain, the protein was detected essentially exclusively in the cell pellet. On the other hand, the myc-HIP-1(Δ22) protein could be detected in both in the supernatant and the cell pellet. Moreover, the myc-HIP-1(Δ22) protein could be immunoprecipitated by anti-Shh antibody when the *HIP* protein was incubated with Shh protein.

While there is presently no evidence to suggest that the wild-type protein is glycosylated, it is formally possible that the HIP proteins may, under certain circumstances, also be modified post-translationally, such as by O-, S- and/or N-linked glycosylation. Potential Asn-glycosylation sites, relative to the mouse *HIP* protein sequence, include Asn99, Asn416, Asn447 and Asn459. Potential attachment sites for proteoglycan-like GAG chains (e.g., heparan sulfate, chondroitin sulfate and the like) include Ser235.

In order to determine, the expression pattern of the various *HIP* clones across species, *in situ* hybridyzation studies were performed in developing embryos of mice, chicken and fish. As described in the Examples below, *HIP* RNA distribution and its temporal expression is consistent with a role of *HIP* polypeptides as downstream targets of *hedgehog* signaling. *In situ* hybridization of mouse embryos indicate that *HIP* RNA is expressed at low levels at sites where *hedgehog* signaling is minimal, i.e. expression of *Shh*, *Ihh* or *Dhh*, is minimal and a dramatic upregulation of *HIP* expression occurs in response to the *hedgehog* upregulation. Firstly, upregulation of *HIP* polypeptides coincides temporarily with *hh* upregulation and its expression occurs opposite to the site of *hh* gene expression. Secondly, ectopic expression of *HIP* (RNA) occurs in response to ectopic expression of *Shh* in the CNS. Furthermore, *HIP* expression is activated in response to the expression of a dominant negative form of cAmp-dependent protein kinase A (PKA), which also activates other *hh* target genes such as *patched*. Furthermore, analysis of null *Dhh*-deficient mutant mice reveals loss of *HIP* expression in the testes, which is the target site for *Dhh* signaling.

Accordingly, certain aspects of the present invention relate to nucleic acids encoding *HIP* polypeptides, the *HIP* polypeptides themselves (including various fragments),

antibodies immunoreactive with *HIP* proteins, and preparations of such compositions. Moreover, the present invention provides diagnostic and therapeutic assays and reagents for detecting and treating disorders involving, for example, aberrant expression (or loss thereof) of *HIP*, *HIP* ligands (particularly *hedgehog* proteins), or signal transducers thereof.

In addition, drug discovery assays are provided for identifying agents which can modulate the biological function of *HIP* proteins, such as by altering the binding of *HIP* molecules to *hedgehog* proteins or other extracellular/matrix factors, or the ability of the bound *HIP* protein to transduce *hedgehog* signals. Such agents can be useful therapeutically to alter the growth, maintenance and/or differentiation of a tissue, particularly a mesodermally-derived tissue, such cartilage, tissue involved in spermatogenesis and tissue derived from dorsal mesoderm; ectodermally-derived tissue, such as tissue derived from the epidermis, neural tube, neural crest, or head mesenchyme; endodermally-derived tissue, such as tissue derived from the primitive gut. Other aspects of the invention are described below or will be apparent to those skilled in the art in light of the present disclosure.

For convenience, certain terms employed in the specification and appended claims are collected here.

The term "*hedgehog*-binding protein" or "*HIP*" polypeptide refers to a family of polypeptides characterized at least in part by being identical or sharing a degree of sequence homology with all or a portion of the a *HIP* polypeptide represented in any of SEQ ID Nos: 5-8. The *HIP* polypeptides can be cloned or purified from any of a number of eukaryotic organisms, especially vertebrates, and particularly mammals. Moreover, other *HIP* polypeptides can be generated according to the present invention, which polypeptides do not ordinarily exist in nature, but rather are generated by non-natural mutagenic techniques.

A number of features of the *HIP* protein have been observed upon inspection. In particular, we have noted that *HIP* sequence encodes a secreted protein having a secretory signal sequence (e.g., a peptidyl portion which causes extracellular secretion of at least a portion of the protein) corresponding to residues 1-15 of SEQ ID No. 5. A membrane-anchoring domain, e.g., in the form of a transmembrane domain, may be provided by residues corresponding to either 357-377 or 680-700 of SEQ ID No: 5.

A "membrane-anchoring" region refers to sequence of amino acids that is capable of retaining the the *HIP* polypeptide at the cell surface.

A "glycosylated" *HIP* polypeptide is an *HIP* polypeptide having a covalent linkage with a glycosyl group (e.g. a derivatized with a carbohydrate). For instance, the *HIP* protein can be glycosylated on an existing residue, or can be mutated to preclude carbohydrate

attachment, or can be mutated to provide new glycosylation sites, such as for N-linked or O-linked glycosylation.

As used herein, the term "vertebrate *hedgehog* protein" refers to vertebrate intercellular signaling molecules related to the *Drosophilia* hedgehog protein. Three of the vertebrate *hedgehog* proteins, *Desert hedgehog* (*Dhh*), *Sonic hedgehog* (*Shh*) and *Indian hedgehog* (*Ihh*), apparently exist in all vertebrates, including amphibians, fish, birds, and mammals. Other members of this family, such as *Banded hedgehog*, *Cephalic hedgehog*, *tiggy-winkle hedgehog*, and *echidna hedgehog* have been so far identified in fish and/or amphibians. Exemplary *hedgehog* polypeptides are described in PCT applications WO96/17924, WO96/16668, WO95/18856.

As used herein, the term "nucleic acid" refers to polynucleotides such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include, as equivalents, analogs of either RNA or DNA made from nucleotide analogs, and, as applicable to the embodiment being described, single (sense or antisense) and double-stranded polynucleotides.

As used herein, the term "gene" or "recombinant gene" refers to a nucleic acid comprising an open reading frame encoding a *HIP* polypeptide, including both exon and (optionally) intron sequences. A "recombinant gene" refers to nucleic acid encoding a *HIP* polypeptide and comprising *HIP*-encoding exon sequences, though it may optionally include intron sequences which are derived from, for example, a chromosomal *HIP* gene or from an unrelated chromosomal gene. Exemplary recombinant genes encoding the subject *HIP* polypeptide are represented in the appended Sequence Listing. The term "intron" refers to a DNA sequence present in a given *HIP* gene which is not translated into protein and is generally found between exons.

As used herein, the term "transfection" means the introduction of a nucleic acid, e.g., an expression vector, into a recipient cell by nucleic acid-mediated gene transfer. "Transformation", as used herein, refers to a process in which a cell's genotype is changed as a result of the cellular uptake of exogenous DNA or RNA, and, for example, the transformed cell expresses a recombinant form of a *HIP* polypeptide or, where anti-sense expression occurs from the transferred gene, the expression of a naturally-occurring form of the *HIP* protein is disrupted.

As used herein, the term "specifically hybridizes" refers to the ability of a nucleic acid probe/primer of the invention to hybridize to at least 15 consecutive nucleotides of a *HIP* gene, such as a *HIP* sequence designated in any one or more of SEQ ID Nos: 1-4 and 9-14, or a sequence complementary thereto, or naturally occurring mutants thereof, such that it has less than 15%, preferably less than 10%, and more preferably less than 5%

background hybridization to a cellular nucleic acid (e.g., mRNA or genomic DNA) encoding a protein other than a *HIP* protein, as defined herein.

An "effective amount" of a *hedgehog* polypeptide, or a bioactive fragment thereof, with respect to the subject method of treatment, refers to an amount of agonist or antagonist in a preparation which, when applied as part of a desired dosage regimen, provides modulation of growth, differentiation or survival of cells, e.g., modulation of spermatogenesis, neuronal differentiation, or skeletogenesis, e.g., osteogenesis, chondrogenesis, or limb patterning.

As used herein, "phenotype" refers to the entire physical, biochemical, and physiological makeup of a cell, e.g., having any one trait or any group of traits.

The terms "induction" or "induce", as relating to the biological activity of a *hedgehog* protein, refers generally to the process or act of causing to occur a specific effect on the phenotype of cell. Such effect can be in the form of causing a change in the phenotype, e.g., differentiation to another cell phenotype, or can be in the form of maintaining the cell in a particular cell, e.g., preventing dedifferentation or promoting survival of a cell.

A "patient" or "subject" to be treated can mean either a human or non-human animal.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of preferred vector is an episome, i.e., a nucleic acid capable of extra-chromosomal replication. Preferred vectors are those capable of autonomous replication and/expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of "plasmids" which refer generally to circular double stranded DNA loops which, in their vector form are not bound to the chromosome. In the present specification, "plasmid" and "vector" are used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors which serve equivalent functions and which become known in the art subsequently hereto.

"Transcriptional regulatory sequence" is a generic term used throughout the specification to refer to DNA sequences, such as initiation signals, enhancers, and promoters, which induce or control transcription of protein coding sequences with which they are operably linked. In preferred embodiments, transcription of a recombinant *HIP* gene is under the control of a promoter sequence (or other transcriptional regulatory sequence) which controls the expression of the recombinant gene in a cell-type in which

expression is intended. It will also be understood that the recombinant gene can be under the control of transcriptional regulatory sequences which are the same or which are different from those sequences which control transcription of the naturally-occurring forms of *HIP* genes.

As used herein, the term "tissue-specific promoter" means a DNA sequence that serves as a promoter, i.e., regulates expression of a selected DNA sequence operably linked to the promoter, and which effects expression of the selected DNA sequence in specific cells of a tissue, such as cells of neuronal or hematopoietic origin. The term also covers so-called "leaky" promoters, which regulate expression of a selected DNA primarily in one tissue, but can cause at least low level expression in other tissues as well.

As used herein, the term "target tissue" refers to connective tissue, cartilage, bone tissue or limb tissue, which is either present in an animal, e.g., a mammal, e.g., a human or is present in in vitro culture, e.g, a cell culture.

As used herein, a "transgenic animal" is any animal, preferably a non-human mammal, bird or an amphibian, in which one or more of the cells of the animal contain heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. This molecule may be integrated within a chromosome, or it may be extrachromosomally replicating DNA. In an exemplary transgenic animal, the transgene causes cells to express a recombinant form of a *HIP* protein, e.g. either agonistic or antagonistic forms. However, transgenic animals in which the recombinant *HIP* gene is silent are also contemplated, as for example, the FLP or CRE recombinase dependent constructs described below. Moreover, "transgenic animal" also includes those recombinant animals in which gene disruption of one or more *HIP* genes is caused by human intervention, including both recombination and antisense techniques.

The "non-human animals" of the invention include vertebrates such as rodents, non-human primates, livestock, avian species, amphibians, reptiles, etc. The term "chimeric animal" is used herein to refer to animals in which the recombinant gene is found, or in which the recombinant is expressed in some but not all cells of the animal. The term "tissue-specific chimeric animal" indicates that a recombinant *HIP* gene is present and/or expressed or disrupted in some tissues but not others.

As used herein, the term "transgene" means a nucleic acid sequence (encoding, e.g., a *HIP* polypeptide, or pending an antisense transcript thereto), which is partly or entirely heterologous, i.e., foreign, to the transgenic animal or cell into which it is introduced, or, is

homologous to an endogenous gene of the transgenic animal or cell into which it is introduced, but which is designed to be inserted, or is inserted, into the animal's genome in such a way as to alter the genome of the cell into which it is inserted (e.g., it is inserted at a location which differs from that of the natural gene or its insertion results in a knockout). A transgene can include one or more transcriptional regulatory sequences and any other nucleic acid, such as introns, that may be necessary for optimal expression of a selected nucleic acid.

As is well known, genes for a particular polypeptide may exist in single or multiple copies within the genome of an individual. Such duplicate genes may be identical or may have certain modifications, including nucleotide substitutions, additions or deletions, which all still code for polypeptides having substantially the same activity. The term "DNA sequence encoding a *HIP* polypeptide" may thus refer to one or more genes within a particular individual. Moreover, certain differences in nucleotide sequences may exist between individuals of the same species, which are called alleles. Such allelic differences may or may not result in differences in amino acid sequence of the encoded polypeptide yet still encode a protein with the same biological activity.

"Homology" and "identity" each refer to sequence similarity between two polypeptide sequences, with identity being a more strict comparison. Homology and identity can each be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same amino acid residue, then the polypeptides can be referred to as identical at that position; when the equivalent site is occupied by the same amino acid (e.g., identical) or a similar amino acid (e.g., similar in steric and/or electronic nature), then the molecules can be refered to as homologous at that position. A percentage of homology or identity between sequences is a function of the number of matching or homologous positions shared by the sequences. An "unrelated" or "non-homologous" sequence shares less than 40 percent identity, though preferably less than 25 percent identity, with a *HIP* sequence of the present invention.

The term "ortholog" refers to genes or proteins which are homologs via speciation, e.g., closely related and assumed to have common descent based on structural and functional considerations. Orthologous proteins function as recognizably the same activity in different species. The term "paralog" refers to genes or proteins which are homologs via gene duplication, e.g., duplicated variants of a gene within a genome. See also. Fritch, WM (1970) *Syst Zool* 19:99-113.

"Cells," "host cells" or "recombinant host cells" are terms used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in

succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A "chimeric protein" or "fusion protein" is a fusion of a first amino acid sequence encoding a *HIP* polypeptide with a second amino acid sequence defining a domain (e.g. polypeptide portion) foreign to and not substantially homologous with any domain of a *HIP* protein. A chimeric protein may present a foreign domain which is found (albeit in a different protein) in an organism which also expresses the first protein, or it may be an "interspecies", "intergenic", etc. fusion of protein structures expressed by different kinds of organisms. In general, a fusion protein can be represented by the general formula X-*HIP*-Y, wherein *HIP* represents a portion of the fusion protein which is derived from a *HIP* protein, and X and Y are, independently, absent or represent amino acid sequences which are not related to a *HIP* sequences in an organism.

As used herein, a "reporter gene construct" is a nucleic acid that includes a "reporter gene" operatively linked to a transcriptional regulatory sequences. Transcription of the reporter gene is controlled by these sequences. The activity of at least one or more of these control sequences is directly or indirectly regulated by a signal transduction pathway involving a phospholipase, e.g., is directly or indirectly regulated by a second messenger produced by the phospholipase activity. The transcriptional regulatory sequences can include a promoter and other regulatory regions, such as enhancer sequences, that modulate the activity of the promoter, or regulatory sequences that modulate the activity or efficiency of the RNA polymerase that recognizes the promoter, or regulatory sequences that are recognized by effector molecules, including those that are specifically induced upon activation of a phospholipase. For example, modulation of the activity of the promoter may be effected by altering the RNA polymerase binding to the promoter region, or, alternatively, by interfering with initiation of transcription or elongation of the mRNA. Such sequences are herein collectively referred to as transcriptional regulatory elements or sequences. In addition, the construct may include sequences of nucleotides that alter the stability or rate of translation of the resulting mRNA in response to second messages, thereby altering the amount of reporter gene product.

As used herein, the terms "transforming growth factor-beta" and "TGF-β" denote a family of structurally related paracrine polypeptides found ubiquitously in vertebrates, and prototypic of a large family of metazoan growth, differentiation, and morphogenesis factors (see, for review, Massaque et al. (1990) *Ann Rev Cell Biol* 6:597-641; and Sporn et al. (1992) *J Cell Biol* 119:1017-1021). Included in this family are the "bone morphogenetic proteins" or "BMPs", which refers to proteins isolated from bone, and fragments thereof and synthetic peptides which are capable of inducing bone deposition alone or when combined

with appropriate cofactors. Preparation of BMPs, such as BMP-1, -2, -3, and -4, is described in, for example, PCT publication WO 88/00205. Wozney (1989) *Growth Fact Res* 1:267-280 describes additional BMP proteins closely related to BMP-2, and which have been designated BMP-5, -6, and -7. PCT publications WO89/09787 and WO89/09788 describe a protein called "OP-1," now known to be BMP-7. Other BMPs are known in the art.

The term "isolated" as also used herein with respect to nucleic acids, such as DNA or RNA, refers to molecules separated from other DNAs, or RNAs, respectively, that are present in the natural source of the macromolecule. For example, an isolated nucleic acid encoding a *HIP* polypeptide preferably includes no more than 10 kilobases (kb) of nucleic acid sequence which naturally immediately flanks the *HIP* gene in genomic DNA, more preferably no more than 5kb of such naturally occurring flanking sequences, and most preferably less than 1.5kb of such naturally occurring flanking sequence. The term isolated as used herein also refers to a nucleic acid or peptide that is substantially free of cellular material, or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. Moreover, an "isolated nucleic acid" is meant to include nucleic acid fragments which are not naturally occurring as fragments and would not be found in the natural state.

As described below, one aspect of the invention pertains to isolated nucleic acids comprising nucleotide sequences encoding *HIP* polypeptides, and/or equivalents of such nucleic acids. The term nucleic acid as used herein is intended to include fragments as equivalents. The term equivalent is understood to include nucleotide sequences encoding functionally equivalent *HIP* polypeptides or functionally equivalent peptides having an activity of a *HIP* protein such as described herein. Equivalent nucleotide sequences will include sequences that differ by one or more nucleotide substitutions, additions or deletions, such as allelic variants; and will, therefore, include sequences that differ from the nucleotide sequence of the *HIP* coding sequences shown in any one or more of SEQ ID Nos: 1-4 and 9-14 due to the degeneracy of the genetic code. Equivalents will also include nucleotide sequences that hybridize under stringent conditions (i.e., equivalent to about 20-27°C below the melting temperature (T_m) of the DNA duplex formed in about 1M salt) to the nucleotide sequences represented in SEQ ID No: 1, 2, 3, 4, 9, 10, 11, 12, 13 or 14. In one embodiment, equivalents will further include nucleic acid sequences derived from and evolutionarily related to, a nucleotide sequences shown in SEQ ID No: 1, SEQ ID No: 2, SEQ ID No: 3 and SEQ ID No: 4.

Moreover, it will be generally appreciated that, under certain circumstances, it may be advantageous to provide homologs of a *HIP* polypeptide which function in a limited capacity as one of either an agonist (e.g., mimics or potentiates a bioactivity of the wild-type *HIP* protein) or an antagonist (e.g., inhibits a bioactivity of the wild-type *HIP* protein), in

order to promote or inhibit only a subset of the biological activities of the naturally-occurring form of the protein. Thus, specific biological effects can be elicited by treatment with a homolog of limited function. For example, truncated forms of the *hedgehog interacting protein*, e.g., soluble fragments of the extracellular domain, can be provided to competitively inhibit ligand (*hedgehog*) binding to the wild-type *HIP* protein.

Homologs of the subject *HIP* protein can be generated by mutagenesis, such as by discrete point mutation(s), or by truncation. For instance, mutation can give rise to homologs which retain substantially the same, or merely a subset, of the biological activity of the *HIP* polypeptide from which it was derived. Alternatively, antagonistic forms of the protein can be generated which are able to inhibit the function of the naturally occurring form of the protein, such as by competitively binding to *hedgehog* proteins and competing with wild-type *HIP*, or binding to other *hedgehog* interacting proteins (such as subunits of a *hedgehog* receptor) to form unresponsive *hedgehog* receptor complexes. Thus, the *HIP* protein and homologs thereof provided by the subject invention may be either positive or negative regulators of cell growth, death and/or differentiation.

In general, polypeptides referred to herein as having an activity of a *HIP* protein (e.g., are "bioactive") are defined as polypeptides which include an amino acid sequence corresponding (e.g., identical or homologous) to all or a portion of the amino acid sequences of the *HIP* protein shown in SEQ ID No: 5, SEQ ID No: 6, SEQ ID No: 7 or SEQ ID No: 8, and which agonize or antagonize all or a portion of the biological/biochemical activities of a naturally occurring *HIP* protein. Examples of such biological activity includes the ability to bind with high affinity hedgehog proteins. The bioactivity of certain embodiments of the subject *HIP* polypeptides can be characterized in terms of an ability to promote differentiation and/or maintenance of cells and tissue from mesodermally-derived tissue, such as tissue derived from dorsal mesoderm; ectodermally-origin, such as tissue derived from the neural tube, neural crest, or head mesenchyme; or endodermally-derived tissue, such as tissue derived from the primitive gut.

Other biological activities of the subject *HIP* proteins are described herein or will be reasonably apparent to those skilled in the art. According to the present invention, a polypeptide has biological activity if it is a specific agonist or antagonist of a naturally-occurring form of a *HIP* protein.

Preferred nucleic acids encode a *HIP* polypeptide comprising an amino acid sequence at least 60%, 70% or 80% homologous, more preferably at least 85% homologous and most preferably at least 95% homologous with an amino acid sequence of a naturally occurring *HIP* protein, e.g., such as represented in SEQ ID No: 5, SEQ ID No: 6, SEQ ID No: 7 or SEQ ID No: 8. Nucleic acids which encode polypeptides at least about 98-99% homology with an amino acid sequence represented in SEQ ID No: 5, SEQ ID No: 6, SEQ

ID No: 7 or SEQ ID No: 8 are of course also within the scope of the invention, as are nucleic acids identical in sequence with the enumerated *HIP* sequence of the Sequence listing. In one embodiment, the nucleic acid is a cDNA encoding a polypeptide having at least one activity of the subject *HIP* polypeptide.

In certain preferred embodiments, the invention features a purified or recombinant *HIP* polypeptide having peptide chain with a molecular weight in the range of 68kd to 88kd, even more preferably in the range of 76kd to 80kd (for a full-length *HIP* protein). It will be understood that certain post-translational modifications, e.g., glycosylation, phosphorylation and the like, can increase the apparent molecular weight of the *HIP* protein relative to the unmodified polypeptide chain, and cleavage of certain sequences, such as pro-sequences, can likewise decrease the apparent molecular weight. Other preferred HIP polypeptides include: a mature *HIP* polypeptide which lacks the signal sequence peptide, e.g., corresponding to residues 16-700 of SEQ ID No: 5, e.g., having a molecular weight of about 76.8kD: a mature, extracellular fragment (soluble) of the receptor, e.g., corresponding to residues 16-356 of SEQ ID No: 5, e.g., having a molecular weight of about 74.4kD; or, e.g., corresponding to residues 16-679 of SEQ ID No: 5, e.g., having a molecular weight of about 38.6kD. In a preferred embodiments, the nucleic acid encodes a HIP polypeptide which includes the hedgehog binding domain. By a "molecular weight of about" it is meant with in about ±5kd.

Another aspect of the invention provides a nucleic acid which hybridizes under high or low stringency conditions to one or more of the nucleic acids represented by SEQ ID Nos: 1-4 and 9-14. Appropriate stringency conditions which promote DNA hybridization, for example, 6.0 x sodium chloride/sodium citrate (SSC) at about 45°C, followed by a wash of 2.0 x SSC at 50°C, are known to those skilled in the art or can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. For example, the salt concentration in the wash step can be selected from a low stringency of about 2.0 x SSC at 50°C to a high stringency of about 0.2 x SSC at 50°C. In addition, the temperature in the wash step can be increased from low stringency conditions at room temperature, about 22°C, to high stringency conditions at about 65°C.

Nucleic acids, having a sequence that differs from the nucleotide sequences shown in SEQ ID No: 1, SEQ ID No: 2, SEQ ID No: 3 or SEQ ID No: 4 due to degeneracy in the genetic code are also within the scope of the invention. Such nucleic acids encode functionally equivalent peptides (i.e., a peptide having a biological activity of a *HIP* polypeptide) but differ in sequence from the sequence shown in the sequence listing due to degeneracy in the genetic code. For example, a number of amino acids are designated by more than one triplet. Codons that specify the same amino acid, or synonyms (for example, CAU and CAC each encode histidine) may result in "silent" mutations which do not affect

the amino acid sequence of a *HIP* polypeptide. However, it is expected that DNA sequence polymorphisms that do lead to changes in the amino acid sequences of the subject *HIP* polypeptides will exist among, for example, humans. One skilled in the art will appreciate that these variations in one or more nucleotides (up to about 3-5% of the nucleotides) of the nucleic acids encoding polypeptides having an activity of a *HIP* polypeptide may exist among individuals of a given species due to natural allelic variation.

As used herein, a *HIP* gene fragment refers to a nucleic acid having fewer nucleotides than the nucleotide sequence encoding the entire mature form of a *HIP* protein yet which (preferably) encodes a polypeptide which retains some biological activity of the full length protein. Fragment sizes contemplated by the present invention include, for example. 5, 10, 25, 50, 75, 100, or 200 amino acids in length. In a preferred embodiment of a truncated receptor, the polypeptide will include all or a sufficient portion of the ligand domain to bind to a *hedgehog* polypeptide.

As indicated by the examples set out below, *HIP* protein-encoding nucleic acids can be obtained from mRNA present in cells of metazoan organisms. It should also be possible to obtain nucleic acids encoding *HIP* polypeptides of the present invention from genomic DNA from both adults and embryos. For example, a gene encoding a *HIP* protein can be cloned from either a cDNA or a genomic library in accordance with protocols described herein. as well as those generally known to persons skilled in the art. A cDNA encoding a *HIP* protein can be obtained by isolating total mRNA from a cell, such as a mammalian cell, e.g. a human cell, as desired. Double stranded cDNAs can be prepared from the total mRNA. and subsequently inserted into a suitable plasmid or bacteriophage vector using any one of a number of known techniques. The gene encoding a *HIP* protein can also be cloned using established polymerase chain reaction techniques in accordance with the nucleotide sequence information provided by the invention. The nucleic acid of the invention can be DNA or RNA. A preferred nucleic acid is a cDNA including a nucleotide sequence represented by any one of SEQ ID No: 1, SEQ ID No: 2, SEQ ID No: 3, SEQ ID No: 4, SEQ ID No: 9, SEQ ID No: 10, or SEQ ID No: 11, SEQ ID No: 12, SEQ ID No: 13 or SEQ ID No: 14.

Another aspect of the invention relates to the use of the isolated nucleic acid in "antisense" therapy. As used herein, "antisense" therapy refers to administration or *in situ* generation of oligonucleotide probes or their derivatives which specifically hybridize (e.g. binds) under cellular conditions, with the cellular mRNA and/or genomic DNA encoding a subject *HIP* protein so as to inhibit expression of that protein, e.g. by inhibiting transcription and/or translation. The binding may be by conventional base pair complementarity, or, for example, in the case of binding to DNA duplexes, through specific interactions in the major groove of the double helix. In general, "antisense" therapy refers to the range of techniques

generally employed in the art, and includes any therapy which relies on specific binding to oligonucleotide sequences.

An antisense construct of the present invention can be delivered, for example, as an expression plasmid which, when transcribed in the cell, produces RNA which is complementary to at least a unique portion of the cellular mRNA which encodes a *HIP* protein. Alternatively, the antisense construct is an oligonucleotide probe which is generated *ex vivo* and which, when introduced into the cell causes inhibition of expression by hybridizing with the mRNA and/or genomic sequences of a *HIP* gene. Such oligonucleotide probes are preferably modified oligonucleotides which are resistant to endogenous nucleases, e.g. exonucleases and/or endonucleases, and are therefore stable *in vivo*. Exemplary nucleic acid molecules for use as antisense oligonucleotides are phosphoramidate, phosphothioate and methylphosphonate analogs of DNA (see also U.S. Patents 5,176,996; 5,264,564; and 5,256,775), or peptide nucleic acids (PNAs). Additionally, general approaches to constructing oligomers useful in antisense therapy have been reviewed, for example, by Van der Krol et al. (1988) Biotechniques 6:958-976; and Stein et al. (1988) Cancer Res 48:2659-2668.

Accordingly, the modified oligomers of the invention are useful in therapeutic, diagnostic, and research contexts. In therapeutic applications, the oligomers are utilized in a manner appropriate for antisense therapy in general. For such therapy, the oligomers of the invention can be formulated for a variety of routes of administration, including systemic and topical or localized administration. Techniques and formulations generally may be found in Remmington's Pharmaceutical Sciences, Meade Publishing Co., Easton, PA. For systemic administration, injection is preferred, including intramuscular, intravenous, intraperitoneal, and subcutaneous. For injection, the oligomers of the invention can be formulated in liquid solutions, preferably in physiologically compatible buffers such as Hank's solution or Ringer's solution. In addition, the oligomers may be formulated in solid form and redissolved or suspended immediately prior to use. Lyophilized forms are also included.

Systemic administration can also be by transmucosal or transdermal means, or the compounds can be administered orally. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration bile salts and fusidic acid derivatives. In addition, detergents may be used to facilitate permeation. Transmucosal administration may be through nasal sprays or using suppositories. For oral administration, the oligomers are formulated into conventional oral administration forms such as capsules, tablets, and tonics. For topical administration, the oligomers of the invention are formulated into ointments, salves, gels, or creams as generally known in the art.

In addition to use in therapy, the oligomers of the invention may be used as diagnostic reagents to detect the presence or absence of the target DNA or RNA sequences to which they specifically bind. Such diagnostic tests are described in further detail below.

Likewise, the antisense constructs of the present invention, by antagonizing the normal biological activity of a *HIP* protein, e.g., by reducing the level of its expression, can be used in the manipulation of tissue, e.g. tissue maintenance, differentiation or growth, both *in vivo* and *ex vivo*.

Furthermore, the anti-sense techniques (e.g. microinjection of antisense molecules, or transfection with plasmids whose transcripts are anti-sense with regard to a *HIP* mRNA or gene sequence) can be used to investigate the role of *HIP* in developmental events, as well as the normal cellular function of *HIP* in adult tissue. Such techniques can be utilized in cell culture, but can also be used in the creation of transgenic animals (described *infra*).

This invention also provides expression vectors containing a nucleic acid encoding a *HIP* polypeptide, operably linked to at least one transcriptional regulatory sequence. Operably linked is intended to mean that the nucleotide sequence is linked to a regulatory sequence in a manner which allows expression of the nucleotide sequence. Regulatory sequences are art-recognized and are selected to direct expression of the subject *HIP* proteins. Accordingly, the term transcriptional regulatory sequence includes promoters, enhancers and other expression control elements. Such regulatory sequences are described in Goeddel; Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). For instance, any of a wide variety of expression control sequences, sequences that control the expression of a DNA sequence when operatively linked to it, may be used in these vectors to express DNA sequences encoding *HIP* polypeptides of this invention. Such useful expression control sequences, include, for example, a viral LTR, such as the LTR of the Moloney murine leukemia virus, the early and late promoters of SV40, adenovirus or cytomegalovirus immediate early promoter, the lac system, the trp system, the TAC or TRC system, T7 promoter whose expression is directed by T7 RNA polymerase, the major operator and promoter regions of phage λ, the control regions for fd coat protein, the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase, e.g., Pho5, the promoters of the yeast α-mating factors, the polyhedron promoter of the baculovirus system and other sequences known to control the expression of genes of prokaryotic or eukaryotic cells or their viruses, and various combinations thereof. It should be understood that the design of the expression vector may depend on such factors as the choice of the host cell to be transformed and/or the type of protein desired to be expressed.

Moreover, the vector's copy number, the ability to control that copy number and the expression of any other proteins encoded by the vector, such as antibiotic markers, should

also be considered. In one embodiment, the expression vector includes a recombinant gene encoding a polypeptide having an agonistic activity of a subject *HIP* polypeptide, or alternatively, encoding a polypeptide which is an antagonistic form of the *HIP* protein. An exemplary *HIP* polypeptide of the present invention is a soluble truncated form of the protein which retains the ligand binding domain, e.g., retains the ability to bind to *hedgehog* polypeptides. Such expression vectors can be used to transfect cells and thereby produce polypeptides, including fusion proteins, encoded by nucleic acids as described herein.

Moreover, the gene constructs of the present invention can also be used as a part of a gene therapy protocol to deliver nucleic acids, e.g., encoding either an agonistic or antagonistic form of a subject *HIP* proteins or an antisense molecule described above. Thus, another aspect of the invention features expression vectors for *in vivo* or *in vitro* transfection and expression of a *HIP* polypeptide or antisense molecule in particular cell types so as to reconstitute the function of, or alternatively, abrogate all or a portion of the biological function of *HIP*-induced transcription in a tissue in which the naturally-occurring form of the protein is misexpressed (or has been disrupted); or to deliver a form of the protein which alters maintenance or differentiation of tissue, or which inhibits neoplastic or hyperplastic proliferation.

Expression constructs of the subject *HIP* polypeptides, as well as antisense constructs, may be administered in any biologically effective carrier, e.g. any formulation or composition capable of effectively delivering the recombinant gene to cells *in vivo*. Approaches include insertion of the subject gene in viral vectors including recombinant retroviruses, adenovirus, adeno-associated virus, and herpes simplex virus-1, or recombinant bacterial or eukaryotic plasmids. Viral vectors transfect cells directly; plasmid DNA can be delivered with the help of, for example, cationic liposomes (lipofectin) or derivatized (e.g. antibody conjugated), polylysine conjugates, gramacidin S, artificial viral envelopes or other such intracellular carriers, as well as direct injection of the gene construct or $CaPO_4$ precipitation carried out *in vivo*. It will be appreciated that because transduction of appropriate target cells represents the critical first step in gene therapy, choice of the particular gene delivery system will depend on such factors as the phenotype of the intended target and the route of administration, e.g. locally or systemically. Furthermore, it will be recognized that the particular gene construct provided for *in vivo* transduction of *HIP* expression are also useful for *in vitro* transduction of cells, such as for use in the *ex vivo* tissue culture systems described below.

A preferred approach for *in vivo* introduction of nucleic acid into a cell is by use of a viral vector containing nucleic acid, e.g. a cDNA encoding the particular *HIP* polypeptide desired. Infection of cells with a viral vector has the advantage that a large proportion of the targeted cells can receive the nucleic acid. Additionally, molecules encoded within the viral

vector, e.g., by a cDNA contained in the viral vector, are expressed efficiently in cells which have taken up viral vector nucleic acid. Retrovirus vectors, adenovirus vectors and adeno-associated virus vectors are exemplary recombinant gene delivery system for the transfer of exogenous genes *in vivo*, particularly into humans. These vectors provide efficient delivery of genes into cells, and the transferred nucleic acids are stably integrated into the chromosomal DNA of the host.

In addition to viral transfer methods, such as those illustrated above, non-viral methods can also be employed to cause expression of a subject *HIP* polypeptide in the tissue of an animal. Most nonviral methods of gene transfer rely on normal mechanisms used by mammalian cells for the uptake and intracellular transport of macromolecules. In preferred embodiments, non-viral gene delivery systems of the present invention rely on endocytic pathways for the uptake of the subject *HIP* polypeptide gene by the targeted cell. Exemplary gene delivery systems of this type include liposomal derived systems, poly-lysine conjugates, and artificial viral envelopes.

In clinical settings, the gene delivery systems for the therapeutic *HIP* gene can be introduced into a patient-animal by any of a number of methods, each of which is familiar in the art. For instance, a pharmaceutical preparation of the gene delivery system can be introduced systemically, e.g. by intravenous injection, and specific transduction of the protein in the target cells occurs predominantly from specificity of transfection provided by the gene delivery vehicle, cell-type or tissue-type expression due to the transcriptional regulatory sequences controlling expression of the receptor gene, or a combination thereof. In other embodiments, initial delivery of the recombinant gene is more limited with introduction into the animal being quite localized. For example, the gene delivery vehicle can be introduced by catheter (see U.S. Patent 5,328,470) or by stereotactic injection (e.g. Chen et al. (1994) PNAS 91: 3054-3057). A *HIP* gene can be delivered in a gene therapy construct by electroporation using techniques described, for example, by Dev et al. ((1994) Cancer Treat Rev 20:105-115).

The pharmaceutical preparation of the gene therapy construct can consist essentially of the gene delivery system in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery system can be produced intact from recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can comprise one or more cells which produce the gene delivery system.

In yet another embodiment, the subject invention provides a "gene activation" construct which, by homologous recombination with a genomic DNA, alters the transcriptional regulatory sequences of an endogenous *HIP* gene. For instance, the gene activation construct can replace the endogenous promoter of a *HIP* gene with a heterologous

promoter. e.g., one which causes constitutive expression of the *HIP* gene or which causes inducible expression of the gene under conditions different from the normal expression pattern of *HIP*. A variety of different formats for the gene activation constructs are available. See, for example, the Transkaryotic Therapies, Inc PCT publications WO93/09222, WO95/31560, WO96/29411, WO95/31560 and WO94/12650.

In preferred embodiments, the nucleotide sequence used as the gene activation construct can be comprised of (1) DNA from some portion of the endogenous *HIP* gene (exon sequence, intron sequence, promoter sequences, etc.) which direct recombination and (2) heterologous transcriptional regulatory sequence(s) which is to be operably linked to the coding sequence for the genomic *HIP* gene upon recombination of the gene activation construct. For use in generating cultures of *HIP* producing cells, the construct may further include a reporter gene to detect the presence of the knockout construct in the cell.

The gene activation construct is inserted into a cell, and integrates with the genomic DNA of the cell in such a position so as to provide the heterologous regulatory sequences in operative association with the native *HIP* gene. Such insertion occurs by homologous recombination, i.e., recombination regions of the activation construct that are homologous to the endogenous *HIP* gene sequence hybridize to the genomic DNA and recombine with the genomic sequences so that the construct is incorporated into the corresponding position of the genomic DNA.

The terms "recombination region" or "targeting sequence" refer to a segment (i.e., a portion) of a gene activation construct having a sequence that is substantially identical to or substantially complementary to a genomic gene sequence, e.g., including 5' flanking sequences of the genomic gene, and can facilitate homologous recombination between the genomic sequence and the targeting transgene construct.

As used herein, the term "replacement region" refers to a portion of a activation construct which becomes integrated into an endogenous chromosomal location following homologous recombination between a recombination region and a genomic sequence.

The heterologous regulatory sequences, e.g., which are provided in the replacement region, can include one or more of a variety elements, including: promoters (such as constitutive or inducible promoters), enhancers, negative regulatory elements, locus control regions, transcription factor binding sites, or combinations thereof. Promoters/enhancers which may be used to control the expression of the targeted gene *in vivo* include, but are not limited to, the cytomegalovirus (CMV) promoter/enhancer (Karasuyama et al., 1989, *J. Exp. Med.*, 169:13), the human β-actin promoter (Gunning et al. (1987) *PNAS* 84:4831-4835), the glucocorticoid-inducible promoter present in the mouse mammary tumor virus long terminal repeat (MMTV LTR) (Klessig et al. (1984) *Mol. Cell Biol.* 4:1354-1362), the long terminal repeat sequences of Moloney murine leukemia virus (MuLV LTR) (Weiss et al.

(1985) *RNA Tumor Viruses*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York), the SV40 early or late region promoter (Bernoist et al. (1981) *Nature* 290:304-310; Templeton et al. (1984) *Mol. Cell Biol.*, 4:817; and Sprague et al. (1983) *J. Virol.*, 45:773), the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (RSV) (Yamamoto et al., 1980, *Cell*, 22:787-797), the herpes simplex virus (HSV) thymidine kinase promoter/enhancer (Wagner et al. (1981) *PNAS* 82:3567-71), and the herpes simplex virus LAT promoter (Wolfe et al. (1992) *Nature Genetics*, 1:379-384).

In still other embodiments, the replacement region merely deletes a negative transcriptional control element of the native gene, e.g., to activate expression, or ablates a positive control element, e.g., to inhibit expression of the targeted gene.

Another aspect of the present invention concerns recombinant forms of the *HIP* proteins. Recombinant polypeptides preferred by the present invention, in addition to native *HIP* proteins, are at least 60% or 70% homologous, more preferably at least 80% homologous and most preferably at least 85% homologous with an amino acid sequence represented by one or more of SEQ ID Nos: 5, SEQ ID No: 6, SEQ ID No: 7 and SEQ ID No: 8. Polypeptides which possess an activity of a *HIP* protein (i.e. either agonistic or antagonistic), and which are at least 90%, more preferably at least 95%, and most preferably at least about 98-99% homologous with SEQ ID No: 5, SEQ ID No: 6, SEQ ID No: 7 and/or SEQ ID No: 8 are also within the scope of the invention. Such polypeptides, as described above, include various truncated forms of the protein.

The term "recombinant *HIP* polypeptide" refers to a polypeptide which is produced by recombinant DNA techniques, wherein generally, DNA encoding a *HIP* polypeptide is inserted into a suitable expression vector which is in turn used to transform a host cell to produce the heterologous protein. Moreover, the phrase "derived from", with respect to a recombinant *HIP* gene, is meant to include within the meaning of "recombinant protein" those proteins having an amino acid sequence of a native *HIP* protein, or an amino acid sequence similar thereto which is generated by mutations including substitutions and deletions (including truncation) of a naturally occurring form of the protein.

The present invention further pertains to recombinant forms of the subject *HIP* polypeptides which are encoded by genes derived from a mammal (e.g. a human), reptile or amphibian and which have amino acid sequences evolutionarily related to the *HIP* protein represented in SEQ ID No: 5, SEQ ID No: 6, SEQ ID No: 7 and SEQ ID No: 8. Such recombinant *HIP* polypeptides preferably are capable of functioning in one of either role of an agonist or antagonist of at least one biological activity of a wild-type ("authentic") *HIP* protein of the appended sequence listing. The term "evolutionarily related to", with respect to amino acid sequences of *HIP* proteins, refers to both polypeptides having amino acid

sequences which have arisen naturally, and also to mutational variants of *HIP* polypeptides which are derived, for example, by combinatorial mutagenesis.

The present invention also provides methods of producing the subject *HIP* polypeptides. For example, a host cell transfected with a nucleic acid vector directing expression of a nucleotide sequence encoding the subject polypeptides can be cultured under appropriate conditions to allow expression of the peptide to occur. If the recombinant protein is not provided with a secretion signal peptide, such as in the case of a GST fuiosn protein, the cells may be harvested, lysed and the protein isolated. A cell culture includes host cells, media and other byproducts. Suitable media for cell culture are well known in the art. The recombinant *HIP* polypeptide can be isolated from cell culture medium, host cells, or both using techniques known in the art for purifying proteins including ion-exchange chromatography, gel filtration chromatography, ultrafiltration, electrophoresis, and immunoaffinity purification with antibodies specific for such peptide. In a preferred embodiment, the recombinant *HIP* polypeptide is a fusion protein containing a domain which facilitates its purification, such as GST fusion protein or poly(His) fusion protein.

This invention also pertains to a host cell transfected to express recombinant forms of the subject *HIP* polypeptides. The host cell may be any eukaryotic or prokaryotic cell. Thus, a nucleotide sequence derived from the cloning of *HIP* proteins, encoding all or a selected portion of a full-length protein, can be used to produce a recombinant form of a *HIP* polypeptide via microbial or eukaryotic cellular processes. Ligating the polynucleotide sequence into a gene construct, such as an expression vector, and transforming or transfecting into hosts, either eukaryotic (yeast, avian, insect or mammalian) or prokaryotic (bacterial cells), are standard procedures used in producing other well-known proteins, e.g. hedgehog proteins, TGFβ proteins, as well as a wide range of receptors. Similar procedures, or modifications thereof, can be employed to prepare recombinant *HIP* polypeptides by microbial means or tissue-culture technology in accord with the subject invention.

The recombinant *HIP* genes can be produced by ligating nucleic acid encoding a *HIP* polypeptide into a vector suitable for expression in either prokaryotic cells, eukaryotic cells, or both. Expression vectors for production of recombinant forms of the subject *HIP* polypeptides include plasmids and other vectors. For instance, suitable vectors for the expression of a *HIP* polypeptide include plasmids of the types: pBR322-derived plasmids, pEMBL-derived plasmids, pEX-derived plasmids, pBTac-derived plasmids and pUC-derived plasmids for expression in prokaryotic cells, such as E. coli.

A number of vectors exist for the expression of recombinant proteins in yeast. For instance, YEP24, YIP5, YEP51, YEP52, pYES2, and YRP17 are cloning and expression vehicles useful in the introduction of genetic constructs into S. cerevisiae (see, for example,

Broach et al. (1983) in Experimental Manipulation of Gene Expression, ed. M. Inouye Academic Press, p. 83, incorporated by reference herein). These vectors can replicate in E. coli due the presence of the pBR322 ori, and in S. cerevisiae due to the replication determinant of the yeast 2 micron plasmid. In addition, drug resistance markers such as ampicillin can be used. In an illustrative embodiment, a *HIP* polypeptide is produced recombinantly utilizing an expression vector generated by sub-cloning the coding sequence of a *HIP* gene represented in SEQ ID No: 1, SEQ ID No: 2, SEQ ID No: 3 or SEQ ID No: 4.

The preferred mammalian expression vectors contain both prokaryotic sequences, to facilitate the propagation of the vector in bacteria, and one or more eukaryotic transcription units that are expressed in eukaryotic cells. The pcDNAI/amp, pcDNAI/neo, pRc/CMV, pSV2gpt, pSV2neo, pSV2-dhfr, pTk2, pRSVneo, pMSG, pSVT7, pko-neo and pHyg derived vectors are examples of mammalian expression vectors suitable for transfection of eukaryotic cells. Some of these vectors are modified with sequences from bacterial plasmids, such as pBR322, to facilitate replication and drug resistance selection in both prokaryotic and eukaryotic cells. Alternatively, derivatives of viruses such as the bovine papillomavirus (BPV-1), or Epstein-Barr virus (pHEBo, pREP-derived and p205) can be used for transient expression of proteins in eukaryotic cells. The various methods employed in the preparation of the plasmids and transformation of host organisms are well known in the art. For other suitable expression systems for both prokaryotic and eukaryotic cells, as well as general recombinant procedures, see Molecular Cloning A Laboratory Manual, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989) Chapters 16 and 17.

In some instances, it may be desirable to express the recombinant *HIP* polypeptide by the use of a baculovirus expression system. Examples of such baculovirus expression systems include pVL-derived vectors (such as pVL1392, pVL1393 and pVL941), pAcUW-derived vectors (such as pAcUW1), and pBlueBac-derived vectors (such as the ß-gal containing pBlueBac III).

When it is desirable to express only a portion of a *HIP* protein, such as a form lacking a portion of the N-terminus, i.e. a truncation mutant which lacks the signal peptide, it may be necessary to add a start codon (ATG) to the oligonucleotide fragment containing the desired sequence to be expressed. It is well known in the art that a methionine at the N-terminal position can be enzymatically cleaved by the use of the enzyme methionine aminopeptidase (MAP). MAP has been cloned from E. coli (Ben-Bassat et al. (1987) J. Bacteriol. 169:751-757) and Salmonella typhimurium and its *in vitro* activity has been demonstrated on recombinant proteins (Miller et al. (1987) PNAS 84:2718-1722). Therefore, removal of an N-terminal methionine, if desired, can be achieved either *in vivo* by expressing *HIP*-derived polypeptides in a host which produces MAP (e.g., E. coli or

CM89 or S. cerevisiae), or *in vitro* by use of purified MAP (e.g., procedure of Miller et al., supra).

Alternatively, the coding sequences for the polypeptide can be incorporated as a part of a fusion gene including a nucleotide sequence encoding a different polypeptide. This type of expression system can be useful under conditions where it is desirable to produce an immunogenic fragment of a *HIP* protein. For example, the VP6 capsid protein of rotavirus can be used as an immunologic carrier protein for portions of the *HIP* polypeptide, either in the monomeric form or in the form of a viral particle. The nucleic acid sequences corresponding to the portion of a subject *HIP* protein to which antibodies are to be raised can be incorporated into a fusion gene construct which includes coding sequences for a late vaccinia virus structural protein to produce a set of recombinant viruses expressing fusion proteins comprising *HIP* epitopes as part of the virion. It has been demonstrated with the use of immunogenic fusion proteins utilizing the Hepatitis B surface antigen fusion proteins that recombinant Hepatitis B virions can be utilized in this role as well. Similarly, chimeric constructs coding for fusion proteins containing a portion of a *HIP* protein and the poliovirus capsid protein can be created to enhance immunogenicity of the set of polypeptide antigens (see, for example, EP Publication No: 0259149; and Evans et al. (1989) Nature 339:385; Huang et al. (1988) J. Virol. 62:3855; and Schlienger et al. (1992) J. Virol. 66:2).

The Multiple Antigen Peptide system for peptide-based immunization can also be utilized to generate an immunogen, wherein a desired portion of a *HIP* polypeptide is obtained directly from organo-chemical synthesis of the peptide onto an oligomeric branching lysine core (see, for example, Posnett et al. (1988) JBC 263:1719 and Nardelli et al. (1992) J. Immunol. 148:914). Antigenic determinants of *HIP* proteins can also be expressed and presented by bacterial cells.

In addition to utilizing fusion proteins to enhance immunogenicity, it is widely appreciated that fusion proteins can also facilitate the expression of proteins, and accordingly, can be used in the expression of the *HIP* polypeptides of the present invention, particularly truncated forms of the *HIP* protein. For example, *HIP* polypeptides can be generated as glutathione-S-transferase (GST-fusion) proteins. Such GST-fusion proteins can enable easy purification of the *HIP* polypeptide, as for example by the use of glutathione-derivatized matrices (see, for example, Current Protocols in Molecular Biology, eds. Ausubel et al. (N.Y.: John Wiley & Sons, 1991)).

In another embodiment, a fusion gene coding for a purification leader sequence, such as a poly-(His)/enterokinase cleavage site sequence at the N-terminus of the desired portion of the recombinant protein, can allow purification of the expressed fusion protein by affinity chromatography using a Ni2+ metal resin. The purification leader sequence can then

be subsequently removed by treatment with enterokinase to provide the purified protein (e.g., see Hochuli et al. (1987) J. Chromatography 411:177; and Janknecht et al. PNAS 88:8972).

Techniques for making fusion genes are known to those skilled in the art. Essentially, the joining of various DNA fragments coding for different polypeptide sequences is performed in accordance with conventional techniques, employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed to generate a chimeric gene sequence (see, for example, Current Protocols in Molecular Biology, eds. Ausubel et al. John Wiley & Sons: 1992).

The *HIP* polypeptides may also be chemically modified to create *HIP* derivatives by forming covalent or aggregate conjugates with other chemical moieties, such as glycosyl groups, lipids, cholesterol, phosphate, acetyl groups and the like. Covalent derivatives of *HIP* proteins can be prepared by linking the chemical moieties to functional groups on amino acid sidechains of the protein or at the N-terminus or at the C-terminus of the polypeptide.

As appropriate, formulations of multimeric *HIP* polypeptides are also provided. The multimers of the soluble forms of the subject *HIP* polypeptides may be produced according to the methods known in the art. In one embodiment, the *HIP* multimers are cross-linked chemically by using known methods which will result in the formation of either dimers or higher multimers of the soluble forms of the *HIP* polypeptides. Another way of producing the multimers of the soluble forms of the *HIP* polypeptides is by recombinant techniques, e.g., by inclusion of hinge regions. This linker can facilitate enhanced flexibility of the chimeric protein allowing the various *HIP* monomeric subunits to freely and (optionally) simultaneously interact with a *HIP* ligand by reducing steric hindrance between the two fragments, as well as allowing appropriate folding of each portion to occur. The linker can be of natural origin, such as a sequence determined to exist in random coil between two domains of a protein. Alternatively, the linker can be of synthetic origin. For instance, the sequence $(Gly_4Ser)_3$ can be used as a synthetic unstructured linker. Linkers of this type are described in Huston et al. (1988) PNAS 85:4879; and U.S. Patent Nos. 5,091,513 and 5,258,498. Naturally occurring unstructured linkers of human origin are preferred as they reduce the risk of immunogenicity.

Each multimer comprises two or more monomers, each comprising the soluble form of a *HIP* polypeptide or a salt or functional derivative thereof. The upper limit for the number of monomers in a multimer is not important and liposomes having many such monomers thereon may be used. Such multimers preferably have 2-5 monomers and more preferably 2 or 3.

The present invention also makes available isolated *HIP* polypeptides which are isolated from, or otherwise substantially free of other cellular proteins, especially receptors and/or other inductive polypeptides which may normally be associated with the *HIP* polypeptide. The term "substantially free of other cellular proteins" (also referred to herein as "contaminating proteins") or "substantially pure or purified preparations" are defined as encompassing preparations of *HIP* polypeptides having less than 20% (by dry weight) contaminating protein, and preferably having less than 5% contaminating protein. Functional forms of the subject polypeptides can be prepared, for the first time, as purified preparations by using a cloned gene as described herein. By "purified", it is meant, when referring to a peptide or DNA or RNA sequence, that the indicated molecule is present in the substantial absence of other biological macromolecules, such as other proteins. The term "purified" as used herein preferably means at least 80% by dry weight, more preferably in the range of 95-99% by weight, and most preferably at least 99.8% by weight, of biological macromolecules of the same type present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 5000, can be present). The term "pure" as used herein preferably has the same numerical limits as "purified" immediately above. "Isolated" and "purified" do not encompass either natural materials in their native state or natural materials that have been separated into components (e.g., in an acrylamide gel) but not obtained either as pure (e.g. lacking contaminating proteins, or chromatography reagents such as denaturing agents and polymers, e.g. acrylamide or agarose) substances or solutions. In preferred embodiments, purified *HIP* preparations will lack any contaminating proteins from the same animal from that *HIP* is normally produced, as can be accomplished by recombinant expression of, for example, a mammalian *HIP* protein in a yeast or bacterial cell.

As described above for recombinant polypeptides, isolated *HIP* polypeptides can include all or a portion of an amino acid sequences corresponding to a *HIP* polypeptide represented in SEQ ID No: 5, SEQ ID No: 6, SEQ ID No: 7 and SEQ ID No: 8 or homologous sequences thereto.

Isolated peptidyl portions of *HIP* proteins can also be obtained by screening peptides recombinantly produced from the corresponding fragment of the nucleic acid encoding such peptides. In addition, fragments can be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry. For example,

a *HIP* polypeptide of the present invention may be arbitrarily divided into fragments of desired length with no overlap of the fragments, or preferably divided into overlapping fragments of a desired length. The fragments can be produced (recombinantly or by chemical synthesis) and tested to identify those peptidyl fragments which can function as either agonists or antagonists of a wild-type (e.g., "authentic") *HIP* protein. For example, Román et al. (1994) *Eur J Biochem* 222:65-73 describe the use of competitive-binding assays using short, overlapping synthetic peptides from larger proteins to identify binding domains.

The recombinant *HIP* polypeptides of the present invention also include homologs of the authentic *HIP* proteins, such as versions of those protein which are resistant to proteolytic cleavage, as for example, due to mutations which alter ubiquitination, prenylation or the like, enzymatic release of the extracellular domain, or other enzymatic targeting associated with the protein.

Modification of the structure of the subject *HIP* polypeptides can be for such purposes as enhancing therapeutic or prophylactic efficacy, stability (e.g., *ex vivo* shelf life and resistance to proteolytic degradation *in vivo*), or post-translational modifications. Such modified peptides, when designed to retain at least one activity of the naturally-occurring form of the protein, or to produce specific antagonists thereof, are considered functional equivalents of the *HIP* polypeptides (though they may be agonistic or antagonistic of the bioactivities of the authentic protein). Such modified peptides can be produced, for instance, by amino acid substitution, deletion, or addition.

For example, it is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid (i.e. isosteric and/or isoelectric mutations) will not have a major effect on the biological activity of the resulting molecule. Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino acids are can be divided into four families: (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine, histidine; (3) nonpolar = alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar = glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. In similar fashion, the amino acid repertoire can be grouped as (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine histidine, (3) aliphatic = glycine, alanine, valine, leucine, isoleucine, serine, threonine, with serine and threonine optionally be grouped separately as aliphatic-hydroxyl; (4) aromatic = phenylalanine, tyrosine, tryptophan; (5) amide = asparagine, glutamine; and (6) sulfur - containing = cysteine and methionine. (see, for example, Biochemistry, 2nd ed., Ed. by L. Stryer, WH Freeman and Co.: 1981). Whether a change in the amino acid sequence of a

peptide results in a functional *HIP* homolog (e.g. functional in the sense that the resulting polypeptide mimics or antagonizes the authentic form) can be readily determined by assessing the ability of the variant peptide to produce a response in cells in a fashion similar to the wild-type protein, or competitively inhibit such a response. Polypeptides in which more than one replacement has taken place can readily be tested in the same manner.

This invention further contemplates a method for generating sets of combinatorial point mutants of the subject *HIP* proteins as well as truncation mutants, and is especially useful for identifying potential variant sequences (e.g. homologs) that are functional in modulating signal transduction and/or ligand binding. The purpose of screening such combinatorial libraries is to generate, for example, novel *HIP* homologs which can act as either agonists or antagonist, or alternatively, possess novel activities all together. To illustrate. *HIP* homologs can be engineered by the present method to provide selective, constitutive activation of *hedgehog* activity, or alternatively, to be dominant negative inhibitors of *HIP*-dependent signal transduction. For instance, mutagenesis can provide *HIP* homologs which are able to bind extracellular ligands yet be unable to bind or signal through intracellular regulatory proteins.

In one aspect of this method, the amino acid sequences for a population of *HIP* homologs from different species or other related proteins are aligned, preferably to promote the highest homology possible. Such a population of variants can include, for example, *HIP* homologs from one or more species. Amino acids which appear at each position of the aligned sequences are selected to create a degenerate set of combinatorial sequences. In a preferred embodiment, the variegated library of *HIP* variants is generated by combinatorial mutagenesis at the nucleic acid level, and is encoded by a variegated gene library. For instance, a mixture of synthetic oligonucleotides can be enzymatically ligated into gene sequences such that the degenerate set of potential *HIP* sequences are expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g. for phage display) containing the set of *HIP* sequences therein.

In an illustrative embodiment, the full-length sequences aligned in Figure 1 are compared in order to generate a degenerate library of potential HIP agonists and antagonists. For instance, a library of HIP polypeptides can be generated to include a degenerate core polypeptide sequence represented by the general formula:

LXFFEGDAKFGEXXXXSGARRRRCLNGXPXXXXXXRXRXXXXXXXXXXXGGXXXXXCXGXY
PRXSCCXXXDXXGLXXXXXKIXSXTNNXECXXLLEEIXCAXCSPHXQXLFXTPEXXXXXX
XXXXLPXLCKDYCKEFFYTCRGHIPGXLQTTADEFCFYYARKDXGLCFPDFPRKQVRGPA
SNYLXXMEXYXKXXXXISRKHKHNCXCXQEVXSGLRQPVXAXHXGDGXXRLFILEKEGYVK
IXXPEGXXXKEPXLDIHKLVQSGIKGGDERGLLSLAFHPNYKKNGKLYVSYTTNQERWAI
GPHDHILRVVEYTVSRKNPXQVDXRTARXFLEVAELHRKHLGGQLLFGPDGFLYXXLGDG
MITLDDMEEMDGLSDFTGSVLRLDVXTDXCXVPYSIPRSNPHFNSTNQPPEXFAHGLHXP
GRCAVDXHPTDXNINLTILCSDSNGKNRSSARILQIIKGRDYESEPSLLEFKPFSXGXLV

GGFVYRGCQSERLYGSYVFGDRNGNFLTLQQXPXTKQWQEKPLCLGXSXSCRGXFSGXXL
GFGEDELGEXYILSSSKSMTQTHNGKLYKIXDPKRPLXPEECXXTXXXAQXLTSXCSRXC
RNGXXTPTGKCCCXXGWEGXFCRXAKCXPACRHGGVCVRPNKCLCKKGYLGPQCEQ
(SEQ ID No. 15)

where each occurence of X is, independently, any (natural) amino acid residue, though more preferably is an amino acid residue (or gap) selected from those residues occuring at the corresponding position in the mouse, human or chicken proteins shown in Figure 1 or a conservative substitution therefor, and even more preferably is an amino acid residue (or gap) selected from those residues occuring at the corresponding position in the mouse, human or chicken proteins shown in Figure 1. As appropriate for the screening assay, the polypeptides of the library can include a secretion signal sequence and/or a C-terminal membrane anchor sequence derived from one of the HIP proteins.

In another embodiment, the degenerate library is based on comparison of the human and mouse sequences, and may include a degenerate core polypeptide sequence represented by the general formula:

LGFFEGDAKFGERXEGSGARRRRCLNGNPPKRLKRRDRRXMSQLELLSGGEXLCGGFYPR
XSCCLXSDSPGLGRLENKIFSXTNNXECXXLLEEIXCAXCSPHSQSLFXXPERXVLXXDX
XLPLLCKDYCKEFFYTCRGHIPGXLQTTADEFCFYYARKDXGLCFPDFPRKQVRGPASNY
LXQMEXYXKVXXISRKHKHNCXCXQEVXSGLRQPVXAXHSGDGSXRLFILEKEGYVKILT
PEGEXFKEPYLDIHKLVQSGIKGGDERGLLSLAFHPNYKKNGKLYVSYTTNQERWAIGPH
DHILRVVEYTVSRKNPHQVDXRTARXFLEVAELHRKHLGGQLLFGPDGFLYIILGDGMIT
LDDMEEMDGLSDFTGSVLRLDVDTDMCNVPYSIPRSNPHFNSTNQPPEVFAHGLHDPGRC
AVDRHPTDININLTILCSDSNGKNRSSARILQIIKGRDYESEPSLLEFKPFSNGPLVGGF
VYRGCQSERLYGSYVFGDRNGNFLTLQQSPVTKQWQEKPLCLGXSXSCRGYFSGHILGFG
EDELGEVYILSSSKSMTQTHNGKLYKIVDPKRPLMPEECRXTVQPAQXLTSXCSRLCRNG
YXTPTGKCCCSPGWEGDFCRXAKCEPACRHGGVCVRPNKCLCKKGYLGPQCEQVDRNXRR
VTR
(SEQ ID No. 16)

where each occurence of X is, independently, any (natural) amino acid residue, though more preferably is an amino acid residue (or gap) selected from those residues occuring at the corresponding position in the mouse or human proteins shown in Figure 1 or a conservative substitution therefor, and even more preferably is an amino acid residue (or gap) selected from those residues occuring at the corresponding position in the mouse or human proteins shown in Figure 1.

There are many ways by which such libraries of potential *HIP* homologs can be generated from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be carried out in an automatic DNA synthesizer, and the synthetic genes then ligated into an appropriate expression vector. The purpose of a degenerate set of genes is to provide, in one mixture, all of the sequences encoding the desired set of potential *HIP* sequences. The synthesis of degenerate oligonucleotides is well known in the art (see for example, Narang, SA (1983) Tetrahedron 39:3; Itakura et al. (1981) Recombinant DNA,

Proc 3rd Cleveland Sympos. Macromolecules, ed. AG Walton, Amsterdam: Elsevier pp273-289; Itakura et al. (1984) Annu. Rev. Biochem. 53:323; Itakura et al. (1984) Science 198:1056; Ike et al. (1983) Nucleic Acid Res. 11:477. Such techniques have been employed in the directed evolution of other proteins (see, for example, Scott et al. (1990) Science 249:386-390; Roberts et al. (1992) PNAS 89:2429-2433; Devlin et al. (1990) Science 249: 404-406; Cwirla et al. (1990) PNAS 87: 6378-6382; as well as U.S. Patents Nos. 5,223,409, 5,198,346, and 5,096,815).

Likewise, a library of coding sequence fragments can be provided for a *HIP* clone in order to generate a variegated population of *HIP* fragments for screening and subsequent selection of bioactive fragments. A variety of techniques are known in the art for generating such libraries, including chemical synthesis. In one embodiment, a library of coding sequence fragments can be generated by (i) treating a double stranded PCR fragment of a *HIP* coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule; (ii) denaturing the double stranded DNA; (iii) renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products; (iv) removing single stranded portions from reformed duplexes by treatment with S1 nuclease; and (v) ligating the resulting fragment library into an expression vector. By this exemplary method, an expression library can be derived which codes for N-terminal, C-terminal and internal fragments of various sizes.

A wide range of techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a certain property. Such techniques will be generally adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of *HIP* homologs. The most widely used techniques for screening large gene libraries typically comprises cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates relatively easy isolation of the vector encoding the gene whose product was detected.

In an exemplary embodiment, a library of *HIP* variants is expressed as a fusion protein on the surface of a viral particle, and the viral particles panned on a ***hedgehog*** matrix. For instance, in the filamentous phage system, foreign peptide sequences can be expressed on the surface of infectious phage, thereby conferring two significant benefits. First, since these phage can be applied to affinity matrices at very high concentrations, a large number of phage can be screened at one time. Second, since each infectious phage displays the combinatorial gene product on its surface, if a particular phage is recovered from an affinity matrix in low yield, the phage can be amplified by another round of infection. The group of almost identical *E. coli* filamentous phages M13, fd., and f1 are

most often used in phage display libraries, as either of the phage gIII or gVIII coat proteins can be used to generate fusion proteins without disrupting the ultimate packaging of the viral particle (Ladner et al. PCT publication WO 90/02909; Garrard et al.. PCT publication WO 92/09690; Marks et al. (1992) *J. Biol. Chem.* 267:16007-16010; Griffiths et al. (1993) *EMBO J* 12:725-734; Clackson et al. (1991) *Nature* 352:624-628; and Barbas et al. (1992) *PNAS* 89:4457-4461). For example, the recombinant phage antibody system (RPAS, Pharmacia Catalog number 27-9400-01) can be easily modified for use in expressing and screening *HIP* combinatorial libraries by panning on a matrix-immobilized *hedgehog* polypeptides to enrich for *HIP* homologs with enhanced ability to bind the ligand.

The invention also provides for reduction of the *HIP* protein to generate mimetics, e.g. peptide or non-peptide agents, which are able to disrupt a biological activity of a *HIP* polypeptide of the present invention, e.g. as inhibitors of protein-protein interactions, such as with ligand proteins. Thus, such mutagenic techniques as described above are also useful to map the determinants of the *HIP* proteins which participate in protein-protein interactions involved in, for example, interaction of the subject *HIP* polypeptide with *hedgehog* polypeptides. Alternatively, a similar system can be used to derive fragments of a *hedgehog* protein which bind to a *HIP* protein and competitively inhibit binding of the full length *hedgehog* protein.

To further illustrate, the critical residues of either a *HIP* protein or a *hedgehog* protein which are involved in molecular recognition of the other can be determined and used to generate *HIP*-derived or *hedgehog*-derived peptidomimetics which competitively inhibit *Hedgehog/HIP* protein interactions. By employing, for example, scanning mutagenesis to map the amino acid residues of a protein which is involved in binding other proteins, peptidomimetic compounds can be generated which mimic those residues which facilitate the interaction. Such mimetics may then be used to interfere with the normal function of a *HIP* protein (or its ligand). For instance, non-hydrolyzable peptide analogs of such residues can be generated using benzodiazepine (e.g., see Freidinger et al. in Peptides: Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), azepine (e.g., see Huffman et al. in Peptides: Chemistry and Biology, G.R. Marshall ed.. ESCOM Publisher: Leiden, Netherlands, 1988), substituted gamma lactam rings (Garvey et al. in Peptides: Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), keto-methylene pseudopeptides (Ewenson et al. (1986) *J Med Chem* 29:295; and Ewenson et al. in Peptides: Structure and Function (Proceedings of the 9th American Peptide Symposium) Pierce Chemical Co. Rockland, IL, 1985), b-turn dipeptide cores (Nagai et al. (1985) *Tetrahedron Lett* 26:647; and Sato et al. (1986) *J Chem Soc* Perkin Trans 1:1231), and b-aminoalcohols (Gordon et al. (1985) *Biochem Biophys Res Commun* 126:419; and Dann et al. (1986) *Biochem Biophys Res Commun* 134:71).

Another aspect of the invention pertains to an antibody specifically reactive with a *HIP* protein. For example, by using immunogens derived from a *HIP* protein, e.g. based on the cDNA sequences, anti-protein/anti-peptide antisera or monoclonal antibodies can be made by standard protocols (See, for example, Antibodies: A Laboratory Manual ed. by Harlow and Lane (Cold Spring Harbor Press: 1988)). A mammal, such as a mouse, a hamster or rabbit can be immunized with an immunogenic form of the peptide (e.g., a *HIP* polypeptide or an antigenic fragment which is capable of eliciting an antibody response). Techniques for conferring immunogenicity on a protein or peptide include conjugation to carriers or other techniques well known in the art. An immunogenic portion of a *HIP* protein can be administered in the presence of adjuvant. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassays can be used with the immunogen as antigen to assess the levels of antibodies. In a preferred embodiment, the subject antibodies are immunospecific for antigenic determinants of a *HIP* protein of a organism, such as a mammal, e.g. antigenic determinants of a protein represented by SEQ ID No: 5, SEQ ID No: 6, SEQ ID No: 7 and SEQ ID No: 8 or closely related homologs (e.g. at least 70% homologous, preferably at least 80% homologous, and more preferably at least 90% homologous). In yet a further preferred embodiment of the present invention, in order to provide, for example, antibodies which are immuno-selective for discrete *HIP* homologs the anti-*HIP* polypeptide antibodies do not substantially cross react (i.e. does not react specifically) with a protein which is, for example, less than 85%, 90% or 95% homologous with the selected *HIP*. By "not substantially cross react", it is meant that the antibody has a binding affinity for a non-homologous protein which is at least one order of magnitude, more preferably at least 2 orders of magnitude, and even more preferably at least 3 orders of magnitude less than the binding affinity of the antibody for the intended target *HIP*.

Following immunization of an animal with an antigenic preparation of a *HIP* polypeptide, anti-*HIP* antisera can be obtained and, if desired, polyclonal anti-*HIP* antibodies isolated from the serum. To produce monoclonal antibodies, antibody-producing cells (lymphocytes) can be harvested from an immunized animal and fused by standard somatic cell fusion procedures with immortalizing cells such as myeloma cells to yield hybridoma cells. Such techniques are well known in the art, an include, for example, the hybridoma technique (originally developed by Kohler and Milstein, (1975) Nature, 256: 495-497), the human B cell hybridoma technique (Kozbar et al., (1983) Immunology Today, 4: 72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al., (1985) Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc. pp. 77-96). Hybridoma cells can be screened immunochemically for production of antibodies specifically reactive with a *HIP* polypeptide of the present invention and monoclonal antibodies isolated from a culture comprising such hybridoma cells.

The term antibody as used herein is intended to include fragments thereof which are also specifically reactive with a *HIP* polypeptide. Antibodies can be fragmented using conventional techniques and the fragments screened for utility in the same manner as described above for whole antibodies. For example, $F(ab)_2$ fragments can be generated by treating antibody with pepsin. The resulting $F(ab)_2$ fragment can be treated to reduce disulfide bridges to produce Fab fragments. The antibody of the present invention is further intended to include bispecific and chimeric molecules having affinity for a *HIP* protein conferred by at least one CDR region of the antibody.

Both monoclonal and polyclonal antibodies (Ab) directed against authentic *HIP* polypeptides, or *HIP* variants, and antibody fragments such as Fab, $F(ab)_2$, Fv and scFv can be used to block the action of a *HIP* protein and allow the study of the role of these proteins in, for example, differentiation of tissue. Experiments of this nature can aid in deciphering the role of *HIP* proteins that may be involved in control of proliferation versus differentiation, e.g., in patterning and tissue formation.

Antibodies which specifically bind *HIP* epitopes can also be used in immunohistochemical staining of tissue samples in order to evaluate the abundance and pattern of expression of each of the subject *HIP* polypeptides. Anti-*HIP* antibodies can be used diagnostically in immuno-precipitation and immuno-blotting to detect and evaluate *HIP* protein levels in tissue as part of a clinical testing procedure. For instance, such measurements can be useful in predictive valuations of the onset or progression of proliferative or differentiative disorders. Likewise, the ability to monitor *HIP* protein levels in an individual can allow determination of the efficacy of a given treatment regimen for an individual afflicted with such a disorder. The level of *HIP* polypeptides may be measured from cells in bodily fluid, such as in samples of cerebral spinal fluid or amniotic fluid, or can be measured in tissue, such as produced by biopsy. Diagnostic assays using anti-*HIP* antibodies can include, for example, immunoassays designed to aid in early diagnosis of a disorder, particularly ones which are manifest at birth. Diagnostic assays using anti-*HIP* polypeptide antibodies can also include immunoassays designed to aid in early diagnosis and phenotyping neoplastic or hyperplastic disorders.

Another application of anti-*HIP* antibodies of the present invention is in the immunological screening of cDNA libraries constructed in expression vectors such as λ gt11, λgt18-23, λZAP, and λORF8. Messenger libraries of this type, having coding sequences inserted in the correct reading frame and orientation, can produce fusion proteins. For instance, λgt11 will produce fusion proteins whose amino termini consist of ß-galactosidase amino acid sequences and whose carboxy termini consist of a foreign polypeptide. Antigenic epitopes of a *HIP* protein, e.g. orthologs of the *HIP* protein from other species, can then be detected with antibodies, as, for example, reacting nitrocellulose

filters lifted from infected plates with anti-*HIP* antibodies. Positive phage detected by this assay can then be isolated from the infected plate. Thus, the presence of *HIP* homologs can be detected and cloned from other animals, as can alternate isoforms (including splicing variants) from humans.

Moreover, the nucleotide sequences determined from the cloning of *HIP* genes from organisms will further allow for the generation of probes and primers designed for use in identifying and/or cloning *HIP* homologs in other cell types, e.g. from other tissues, as well as *HIP* homologs from other organisms. For instance, the present invention also provides a probe/primer comprising a substantially purified oligonucleotide, which oligonucleotide comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least 15 consecutive nucleotides of sense or anti-sense sequence selected from the group consisting of SEQ ID No: 1, SEQ ID No: 2, SEQ ID No: 3 or SEQ ID No: 4 or naturally occurring mutants thereof. For instance, primers based on the nucleic acid represented in SEQ ID No: 1, SEQ ID No: 2, SEQ ID No: 3 or SEQ ID No: 4, can be used in PCR reactions to clone *HIP* homologs. Likewise, probes based on the subject *HIP* sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto and able to be detected, e.g. the label group is selected from amongst radioisotopes, fluorescent compounds, enzymes, and enzyme co-factors.

Such probes can also be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a *HIP* protein, such as by measuring a level of a *HIP*-encoding nucleic acid in a sample of cells from a patient-animal; e.g. detecting *HIP* mRNA levels or determining whether a genomic *HIP* gene has been mutated or deleted.

To illustrate, nucleotide probes can be generated from the subject *HIP* genes which facilitate histological screening of intact tissue and tissue samples for the presence (or absence) of *HIP*-encoding transcripts. Similar to the diagnostic uses of anti-*HIP* antibodies, the use of probes directed to *HIP* messages, or to genomic *HIP* sequences, can be used for both predictive and therapeutic evaluation of allelic mutations which might be manifest in, for example, degenerative disorders marked by loss of particular cell-types, apoptosis, neoplastic and/or hyperplastic disorders (e.g. unwanted cell growth) or abnormal differentiation of tissue. Used in conjunction with immunoassays as described above, the oligonucleotide probes can help facilitate the determination of the molecular basis for a developmental disorder which may involve some abnormality associated with expression (or lack thereof) of a *HIP* protein. For instance, variation in polypeptide synthesis can be differentiated from a mutation in a coding sequence.

Accordingly, the present method provides a method for determining if a subject is at risk for a disorder characterized by aberrant apoptosis, cell proliferation and/or

differentiation. In preferred embodiments. method can be generally characterized as comprising detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of (i) an alteration affecting the integrity of a gene encoding a *HIP*-protein, or (ii) the mis-expression of the *HIP* gene. To illustrate, such genetic lesions can be detected by ascertaining the existence of at least one of (i) a deletion of one or more nucleotides from a *HIP* gene, (ii) an addition of one or more nucleotides to a *HIP* gene, (iii) a substitution of one or more nucleotides of a *HIP* gene, (iv) a gross chromosomal rearrangement of a *HIP* gene, (v) a gross alteration in the level of a messenger RNA transcript of a *HIP* gene, (vii) aberrant modification of a *HIP* gene, such as of the methylation pattern of the genomic DNA, (vii) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a *HIP* gene, (viii) a non-wild type level of a *HIP*-protein, and (ix) inappropriate post-translational modification of a *HIP*-protein. As set out below, the present invention provides a large number of assay techniques for detecting lesions in a *HIP* gene. and importantly, provides the ability to discern between different molecular causes underlying *HIP*-dependent aberrant cell growth, proliferation and/or differentiation.

In an exemplary embodiment, there is provided a nucleic acid composition comprising a (purified) oligonucleotide probe including a region of nucleotide sequence which is capable of hybridizing to a sense or antisense sequence of a *HIP* gene, such as represented by any one of SEQ ID Nos: 1-4 and 9-14, or naturally occurring mutants thereof, or 5' or 3' flanking sequences or intronic sequences naturally associated with the subject *HIP* genes or naturally occurring mutants thereof. The nucleic acid of a cell is rendered accessible for hybridization, the probe is exposed to nucleic acid of the sample, and the hybridization of the probe to the sample nucleic acid is detected. Such techniques can be used to detect lesions at either the genomic or mRNA level, including deletions, substitutions, etc., as well as to determine mRNA transcript levels.

In certain embodiments, detection of the lesion comprises utilizing the probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran et al. (1988) Science 241:1077-1080; and Nakazawa et al. (1944) PNAS 91:360-364), the later of which can be particularly useful for detecting point mutations in the *HIP* gene. In a merely illustrative embodiment, the method includes the steps of (i) collecting a sample of cells from a patient, (ii) isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, (iii) contacting the nucleic acid sample with one or more primers which specifically hybridize to a *HIP* gene under conditions such that hybridization and amplification of the *HIP* gene (if present) occurs, and (iv) detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample.

In still another embodiment, the level of a *HIP*-protein can be detected by immunoassay. For instance, the cells of a biopsy sample can be lysed, and the level of a *HIP*-protein present in the cell can be quantitated by standard immunoassay techniques. In yet another exemplary embodiment, aberrant methylation patterns of a *HIP* gene can be detected by digesting genomic DNA from a patient sample with one or more restriction endonucleases that are sensitive to methylation and for which recognition sites exist in the *HIP* gene (including in the flanking and intronic sequences). See, for example, Buiting et al. (1994) Human Mol Genet 3:893-895. Digested DNA is separated by gel electrophoresis, and hybridized with probes derived from, for example, genomic or cDNA sequences. The methylation status of the *HIP* gene can be determined by comparison of the restriction pattern generated from the sample DNA with that for a standard of known methylation.

In still other embodiments, the ligand binding domain of the *HIP* receptor can be used to quantitatively detect the level of *HIP* ligands, e.g., *hedgehog* proteins. To illustrate, a soluble form of the *HIP* protein can be generated which retains *hedgehog* binding activity. Samples of bodily fluid(s), e.g., plasma, serum, lymph, marrow, cerebral/spinal fluid, urine and the like can be contacted with the receptor under conditions wherein ligand/receptor binding can occur, and the level of ligand/receptor complexes formed can be detected by any of a variety of techniques known in the art. For example, competitive binding assays using standardized samples of *hedgehog* proteins can be used to quantitate the amount of analyte bound from the fluid sample.

In yet other embodiments, such *HIP* receptors can be used to detect the presence of a *HIP* ligand on a cell surface. For instance, the *HIP* protein can be contacted with cells from a biopsy, and the ability of the *HIP* protein to decorate certain cells of the sample is ascertained. The binding of the *HIP* protein to cell populations of the sample can be detected, for example, by the use of antibodies against the *HIP* protein, or by detection of a label associated with the *HIP* protein. In the case of the latter, the *HIP* protein can be labeled, for example, by chemical modification or as a fusion protein. Exemplary labels include radioisotopes, fluorescent compounds, enzyme co-factors, which can be added by chemical modification of the protein, and epitope tags such as myc, pFLAG and the like, or enzymatic activities such as GST or alkaline phosphatase which can be added either by chemical modification or by generation of a fusion protein.

Furthermore, the present invention also contemplates the detection of soluble forms of the *HIP* receptor in bodily fluid samples. As described in the art, e.g., see Diez-Ruiz et al. (1995) *Eur J Haematol* 54:1-8 and Owen-Schaub et al. (1995) *Cancer Lett* 94:1-8, [describing CNTF receptors] in certain instances soluble forms of receptors are believed to play a role as modulators of the biological function of their cognate ligands in an agonist/antagonist pattern. In various pathologic states, the production and release of

soluble *HIP* proteins may mediate host response and determine the course and outcome of disease by interacting with *HIP* ligands and competing with cell surface receptors. The determination of soluble *HIP* receptors in body fluids is a new tool to gain information about various disease states, and may be of prognostic value to a clinician. For example, the level of soluble *HIP* protein in a body fluid may give useful information for monitoring, *inter alia*, neurological disorders as well as in the treatment of neoplastic or hyperplastic transformations of ectodermal, mesodermal or endodermal origin.

The level of soluble receptor present in a given sample can be quantitated, in light of the present disclosure, using known procedures and techniques. For example, antibodies immunoselective for the ligand binding domain of the *HIP* protein can be used to detect and quantify its presence in a sample, e.g., by well-known immunoassay techniques. Alternatively, a labeled ligand of the receptor can be used to detect the presence of the receptor in the fluid sample.

A number of techniques exist in the art for now identifying additional ligands to the *HIP* receptor. For instance, expression cloning can be carried out on a cDNA or genomic library by isolating cells which are decorated with a labeled form of the receptor. In a preferred embodiment, the technique uses the *HIP* receptor in an *in situ* assay for detecting *HIP* ligands in tissue samples and whole organisms. In general, the RAP-*in situ* assay described below (for Receptor Affinity Probe) of Flanagan and Leder (see PCT publications WO 92/06220; and also Cheng et al. (1994) *Cell* 79:157-168) involves the use of an expression cloning system whereby a *HIP* ligand is scored on the basis of binding to a *HIP*/alkaline phosphatase fusion protein. In general, the method comprises (i) providing a hybrid molecule (the affinity probe) including the *HIP* receptor, or at least the ligand binding domain thereof, covalently bonded to an enzymatically active tag, preferably for which chromogenic substrates exist, (ii) contacting the tissue or organism with the affinity probe to form complexes between the probe and a cognate ligand in the sample, removing unbound probe, and (iii) detecting the affinity complex using a chromogenic substrate for the enzymatic activity associated with the affinity probe.

This method, unlike other prior art methods which are carried out only on dispersed cell cultures, provides a means for probing non-dispersed and wholemount tissue and animal samples. The method can be used, in addition to facilitating the cloning of *HIP* ligands, also for detecting patterns of expression for particular ligands of the *HIP* receptor, for measuring the affinity of receptor/ligand interactions in tissue samples, as well as for generating drug screening assays in tissue samples. Moreover, the affinity probe can also be used in diagnostic screening to determine whether a *HIP* ligand is misexpressed.

In yet another aspect of the invention, the subject *HIP* polypeptides can be used to generate a "two hybrid" assay or an "interaction trap" assay (see, for example, U.S. Patent

No. 5.283,317; Zervos et al. (1993) Cell 72:223-232; Madura et al. (1993) J Biol Chem 268:12046-12054; Bartel et al. (1993) Biotechniques 14:920-924; Iwabuchi et al. (1993) Oncogene 8:1693-1696; and Brent WO94/10300), for isolating coding sequences for other proteins which bind *HIP*s ("*HIP*-binding proteins" or "*HIP*-bp").

Briefly, the interaction trap relies on reconstituting *in vivo* a functional transcriptional activator protein from two separate fusion proteins. In particular, the method makes use of chimeric genes which express hybrid proteins. To illustrate, a first hybrid gene comprises the coding sequence for a DNA-binding domain of a transcriptional activator fused in frame to the coding sequence for a *HIP* polypeptide. The second hybrid protein encodes a transcriptional activation domain fused in frame to a sample gene from a cDNA library. If the bait and sample hybrid proteins are able to interact, e.g., form a *HIP*-dependent complex, they bring into close proximity the two domains of the transcriptional activator. This proximity is sufficient to cause transcription of a reporter gene which is operably linked to a transcriptional regulatory site responsive to the transcriptional activator, and expression of the reporter gene can be detected and used to score for the interaction of the *HIP* and sample proteins.

Furthermore, by making available purified and recombinant *HIP* polypeptides, the present invention facilitates the development of assays which can be used to screen for drugs which are either agonists or antagonists of the normal cellular function of the subject *HIP* proteins, or of their role in the pathogenesis of cellular maintenance, differentiation and/or proliferation and disorders related thereto. In a general sense, the assay evaluates the ability of a compound to modulate binding between a *HIP* polypeptide and a molecule, e.g., a ligand such as a *hedgehog* protein, that interacts with the *HIP* polypeptide. Exemplary compounds which can be screened against such *HIP*-mediated interactions include peptides, nucleic acids, carbohydrates, small organic molecules, and natural product extract libraries, such as isolated from animals, plants, fungus and/or microbes.

In many drug screening programs which test libraries of compounds and natural extracts, high throughput assays are desirable in order to maximize the number of compounds surveyed in a given period of time. Assays which are performed in cell-free systems, such as may be derived with purified or semi-purified proteins, are often preferred as "primary" screens in that they can be generated to permit rapid development and relatively easy detection of an alteration in a molecular target which is mediated by a test compound. Moreover, the effects of cellular toxicity and/or bioavailability of the test compound can be generally ignored in the *in vitro* system, the assay instead being focused primarily on the effect of the drug on the molecular target as may be manifest in an alteration of binding affinity with a ligand. Accordingly, in an exemplary screening assay of the present invention, a reaction mixture is generated to include a *HIP* polypeptide,

compound(s) of interest, and a "target molecule", e.g., a protein, which interacts with the *HIP* polypeptide. Exemplary target molecules include ligands, such as *hedgehog* proteins, as well as other peptide and non-peptide interacting molecules. Detection and quantification of interaction of the *HIP* polypeptide with the target molecule provides a means for determining a compound's efficacy at inhibiting (or potentiating) interaction between the *HIP* and the target molecule. The efficacy of the compound can be assessed by generating dose response curves from data obtained using various concentrations of the test compound. Moreover, a control assay can also be performed to provide a baseline for comparison. In the control assay, interaction of the *HIP* polypeptide and target molecule is quantitated in the absence of the test compound.

Interaction between the *HIP* polypeptide and the target molecule may be detected by a variety of techniques. Modulation of the formation of complexes can be quantitated using, for example, detectably labeled proteins such as radiolabeled, fluorescently labeled, or enzymatically labeled *HIP* polypeptides, by immunoassay, by chromatographic detection, or by detecting the intrinsic activity of the acetylase.

Typically, it will be desirable to immobilize either *HIP* or the target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of *HIP* to the target molecule, in the presence and absence of a candidate agent, can be accomplished in any vessel suitable for containing the reactants. Examples include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase/*HIP* (GST/*HIP*) fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates, e.g. an ^{35}S-labeled, and the test compound, and the mixture incubated under conditions conducive to complex formation, e.g. at physiological conditions for salt and pH, though slightly more stringent conditions may be desired. Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly (e.g. beads placed in scintillant), or in the supernatant after the complexes are subsequently dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of target molecule found in the bead fraction quantitated from the gel using standard electrophoretic techniques.

Other techniques for immobilizing proteins and other molecules on matrices are also available for use in the subject assay. For instance, either *HIP* or target molecule can be immobilized utilizing conjugation of biotin and streptavidin. For instance, biotinylated *HIP* molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques

well known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with *HIP*, but which do not interfere with the interaction between the *HIP* and target molecule, can be derivatized to the wells of the plate, and *HIP* trapped in the wells by antibody conjugation. As above, preparations of an target molecule and a test compound are incubated in the *HIP*-presenting wells of the plate, and the amount of complex trapped in the well can be quantitated. Exemplary methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the target molecule, or which are reactive with *HIP* protein and compete with the target molecule; as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule, either intrinsic or extrinsic activity. In the instance of the latter, the enzyme can be chemically conjugated or provided as a fusion protein with the target molecule. To illustrate, the target molecule can be chemically cross-linked or genetically fused (if it is a polypeptide) with horseradish peroxidase, and the amount of polypeptide trapped in the complex can be assessed with a chromogenic substrate of the enzyme, e.g. 3,3'-diamino-benzadine terahydrochloride or 4-chloro-1-napthol. Likewise, a fusion protein comprising the polypeptide and glutathione-S-transferase can be provided, and complex formation quantitated by detecting the GST activity using 1-chloro-2,4-dinitrobenzene (Habig et al (1974) J Biol Chem 249:7130).

For processes which rely on immunodetection for quantitating proteins trapped in the complex, antibodies against the protein, such as anti-*HIP* antibodies, can be used. Alternatively, the protein to be detected in the complex can be "epitope tagged" in the form of a fusion protein which includes, in addition to the *HIP* sequence, a second polypeptide for which antibodies are readily available (e.g. from commercial sources). For instance, the GST fusion proteins described above can also be used for quantification of binding using antibodies against the GST moiety. Other useful epitope tags include myc-epitopes (e.g., see Ellison et al. (1991) J Biol Chem 266:21150-21157) which includes a 10-residue sequence from c-myc, as well as the pFLAG system (International Biotechnologies, Inc.) or the pEZZ-protein A system (Pharamacia, NJ).

An exemplary drug screening assay of the present invention includes the steps of (a) forming a reaction mixture including: (i) a *hedgehog* polypeptide, (ii) a *HIP* polypeptide, and (iii) a test compound; and (b) detecting interaction of the *hedgehog* and *HIP* polypeptides. A statistically significant change (potentiation or inhibition) in the interaction of the *hedgehog* and *HIP* polypeptides in the presence of the test compound, relative to the interaction in the absence of the test compound, indicates a potential agonist (mimetic or potentiator) or antagonist (inhibitor) of *hedgehog* bioactivity for the test compound. The reaction mixture can be a cell-free protein preparation, e.g., a reconsistuted protein mixture

or a cell lysate, or it can be a recombinant cell including a heterologous nucleic acid recombinantly expressing the *HIP* polypeptide.

Where the *HIP* polypeptide participates as part of an oligomeric complex forming a *hedgehog* receptor, e.g., which complex includes other protein subunits, the cell-free system can be, e.g., a cell membrane preparation, a reconstituted protein mixture, or a liposome reconstituting the receptor subunots as a *hedgehog* receptor. Alternatively, liposomal preparations using reconstituted *Hip* protein can be utilized. For instance, the protein subunits of a *hedgehog* receptor complex can be purified from detergent extracts from both authentic and recombinant origins can be reconstituted in in artificial lipid vesicles (e.g. phosphatidylcholine liposomes) or in cell membrane-derived vesicles (see, for example, Bear et al. (1992) *Cell* 68:809-818; Newton et al. (1983) *Biochemistry* 22:6110-6117; and Reber et al. (1987) *J Biol Chem* 262:11369-11374). The lamellar structure and size of the resulting liposomes can be characterized using electron microscopy. External orientation of the *HIP* protein in the reconstituted membranes can be demonstrated, for example, by immunoelectron microscopy. The interaction of a *hedgehog* protein with liposomes containing such *HIP* complexes and liposomes without the protein, in the presence of candidate agents, can be compared in order to identify potential modulators of the *hedgehog-HIP* polypeptide interaction.

In yet another embodiment, the drug screening assay is derived to include a whole cell expressing a *HIP* polypeptide. The ability of a test agent to alter the activity of the *HIP* protein can be detected by analysis of the recombinant cell. For example, agonists and antagonists of the *HIP* biological activity can by detected by scoring for alterations in growth or differentiation (phenotype) of the cell. General techniques for detecting each are well known, and will vary with respect to the source of the particular reagent cell utilized in any given assay. For the cell-based assays, the recombinant cell is preferably a metazoan cell, e.g., a mammalian cell, e.g., an insect cell, e.g., a xenopus cell, e.g., an oocyte. In other embodiments, the *hedgehog* receptor can be reconsituted in a yeast cell.

In an exemplary embodiment, a cell which expresses the *HIP* receptor, e.g, whether endogenous or heterologous, can be contacted with a ligand of the *HIP* receptor, e.g., a *hedgehog* protein, which is capable of inducing signal transduction from the receptor, and the resulting signaling detected either at various points in the pathway, or on the basis of a phenotypic change to the reagent cell. In one embodiment, the reagent cell is contacted with antibody which causes cross-linking of the receptor, and the signal cascade induced by that cross-linking is subsequently detected. A test compound which modulates that pathway, e.g., potentiates or inhibits, can be detected by comparison with control experiments which either lack the receptor or lack the test compound. For example, visual

inspection of the morphology of the reagent cell can be used to determine whether the biological activity of the targeted *HIP* protein has been affected by the added agent.

In addition to morphological studies, change(s) in the level of an intracellular second messenger responsive to signaling by the *HIP* polypeptide can be detected. For example, in various embodiments the assay may assess the ability of test agent to cause changes in phophorylation patterns, adenylate cyclase activity (cAMP production), GTP hydrolysis, calcium mobilization, and/or phospholipid hydrolysis (IP_3, DAG production) upon receptor stimulation. By detecting changes in intracellular signals, such as alterations in second messengers or gene expression, in cells contacted with a *hedgehog* polypeptide, candidate agonists and antagonists to *HIP*-dependent *hedgehog* signaling can be identified.

The transduction of certain intracellular signals can be initiated by the specific interaction of an *hh* polypeptide with *HIP* protein, while other signals can be indirecetly altered by that iteraction. In Drosophila, and presumptively in vertebrate cells as well, a number of gene products, including *HIP*, *patched*, the transcription factor *cubitus interruptus* (ci), the serine/threonine kinase *fused* (fu) and the gene products of *costal-2*, *smoothened* and *suppressor of fused*, have been implicated as putative components of *hedgehog*-dependent signal transduction pathways. The recent cloning of vertebrate homologs of the drosophila genes suggests that the *hedgehog* signaling pathway is highly conserved from drosophila to vertebrate species. The activity of each of these proteins can be detected directly (such as the kinase activity of *fused*, or can detected indirectly by monitoring the level of second messangers produced downstream in the signal pathway.

To further illustrate, recent studies have implicated protein kinase A (PKA) as a possible component of *hedgehog* signaling in drosophila and vertebrate organisms (Hammerschmidt et al. (1996) *Genes & Dev* 10:647). High PKA activity has been shown to antagonize *hedgehog* signaling in these systems. Although it is unclear whether PKA acts directly downstream or in parallel with *hedgehog* signaling, it is possible that *hedgehog* signaling occurring through a *HIP* protein effects inhibition of PKA activity. Thus, detection of PKA activity provides a potential readout for the instant assays.

Binding of *hedgehog* to *HIP* proteins may stimulate the activity of phospholipases. Inositol lipids can be extracted and analyzed using standard lipid extraction techniques. Water soluble derivatives of all three inositol lipids (IP_1, IP_2, IP_3) can also be quantitated using radiolabelling techniques or HPLC.

The mobilization of intracellular calcium or the influx of calcium from outside the cell may be a response to *hedgehog* stimulation or lack there of. Calcium flux in the reagent cell can be measured using standard techniques. The choice of the appropriate calcium indicator, fluorescent, bioluminescent, metallochromic, or Ca^{++}-sensitive microelectrodes depends on the cell type and the magnitude and time constant of the event under study

(Borle (1990) *Environ Health Perspect* 84:45-56). As an exemplary method of Ca^{++} detection. cells could be loaded with the Ca^{++}sensitive fluorescent dye fura-2 or indo-1. using standard methods. and any change in Ca^{++} measured using a fluorometer.

In certain embodiments of the assay, it may be desirable to screen for changes in cellular phosphorylation. As an example, the drosophila gene *fused* (fu) which encodes a serine/threonine kinase has been identified as a potential downstream target in *hedgehog* signaling. (Preat et al., 1990 *Nature* 347, 87-89; Therond et al. 1993, *Mech. Dev.* 44. 65-80). The ability of compounds to modulate serine/threonine kinase activation could be screened using colony immunoblotting (Lyons and Nelson (1984) *PNAS* 81:7426-7430) using antibodies against phosphorylated serine or threonine residues. Reagents for performing such assays are commercially available, for example, phosphoserine and phosphothreonine specific antibodies which measure increases in phosphorylation of those residues can be purchased from comercial sources.

The interaction of a *hedgehog* protein with a *HIP* protein may set in motion a cascade involving the activation and inhibition of downstream effectors, the ultimate consequence of which is, in some instances, a detectable change in the transcription or translation of a gene. Potential transcriptional targets of *HIP*-dependent *hedgehog* signaling include the *HIP* gene itself, the *patched* gene (Hidalgo and Ingham (1990) *Development* 110, 291-301 ;Marigo et al. (1996) *Development* 122:1225-1233), and the vertebrate homologs of the drosophila cubitus interruptus (ci) gene. the *GLI* genes (Hui et al. (1994) *Dev Biol* 162:402-413). *Patched* gene expression has been shown to be induced in cells of the limb bud and the neural plate that are responsive to *Shh*. (Marigo et al. (1996) *PNAS*, in press; Marigo et al., *supra*). The *GLI* genes encode putative transcription factors having zinc finger DNA binding domains (Orenic et al. (1990) *Genes & Dev* 4:1053-1067; Kinzler et al. (1990) *Mol Cell Biol* 10:634-642). Transcription of the *GLI* gene has been reported to be upregulated in response to *hedgehog* in limb buds, while transcription of the *GLI3* gene is downregulated in response to *hedgehog* induction (Marigo et al. (1996) *Development* 122:1225-1233). By selecting transcriptional regulatory sequences from such target genes, e.g. from *Hip* or *GLI* genes, that are responsible for the up- or down-regulation of these genes in response to *hedgehog* induction, and operatively linking such promoters to a reporter gene, the present invention provides a transcription based assay which is sensitive to the ability of a specific test compound to influence *hedgehog* signalling pathways.

In an exemplary embodiment, the step of detecting interaction of the *hedgehog* and *HIP* polypeptides comprises detecting, in a cell-based assay, change(s) in the level of expression of a gene controlled by a transcriptional regulatory sequence responsive to signaling by the *HIP* polypeptide. Reporter gene based assays of this invention measure the end stage of the above described cascade of events, e.g., transcriptional modulation.

Accordingly, in practicing one embodiment of the assay, a reporter gene construct is inserted into the reagent cell in order to generate a detection signal dependent on *hedgehog* signaling. Expression of the reporter gene, thus, provides a valuable screening tool for the development of compounds that act as agonists or antagonists of *HIP*-dependent *hedgehog* induction.

In practicing one embodiment of the assay, a reporter gene construct is inserted into the reagent cell in order to generate a detection signal dependent on second messengers generated by *HIP*-dependent induction with a *hedgehog* protein. Typically, the reporter gene construct will include a reporter gene in operative linkage with one or more transcriptional regulatory elements responsive to the *hedgehog* activity, with the level of expression of the reporter gene providing the *hedgehog*-dependent detection signal. The amount of transcription from the reporter gene may be measured using any method known to those of skill in the art to be suitable. For example, mRNA expression from the reporter gene may be detected using RNAse protection or RNA-based PCR, or the protein product of the reporter gene may be identified by a characteristic stain or an intrinsic activity. The amount of expression from the reporter gene is then compared to the amount of expression in either the same cell in the absence of the test compound or it may be compared with the amount of transcription in a substantially identical cell that lacks the target receptor protein. Any statistically or otherwise significant difference in the amount of transcription indicates that the test compound has in some manner altered the inductive activity of the *hedgehog* protein.

As described in further detail below, in preferred embodiments the gene product of the reporter is detected by an intrinsic activity associated with that product. For instance, the reporter gene may encode a gene product that, by enzymatic activity, gives rise to a detection signal based on color, fluorescence, or luminescence. In other preferred embodiments, the reporter or marker gene provides a selective growth advantage, e.g., the reporter gene may enhance cell viability, relieve a cell nutritional requirement, and/or provide resistance to a drug. Many reporter genes are known to those of skill in the art and others may be identified or synthesized by methods known to those of skill in the art. A reporter gene includes any gene that expresses a detectable gene product, which may be RNA or protein.

Preferred reporter genes are those that are readily detectable. The reporter gene may also be included in the construct in the form of a fusion gene with a gene that includes desired transcriptional regulatory sequences or exhibits other desirable properties. Examples of reporter genes include, but are not limited to CAT (chloramphenicol acetyl transferase) (Alton and Vapnek (1979), Nature 282: 864-869) luciferase, and other enzyme detection systems, such as beta-galactosidase; firefly luciferase (deWet et al. (1987), Mol.

Cell. Biol. 7:725-737); bacterial luciferase (Engebrecht and Silverman (1984), PNAS 1: 4154-4158; Baldwin et al. (1984), Biochemistry 23: 3663-3667); alkaline phosphatase (Toh et al. (1989) Eur. J. Biochem. 182: 231-238, Hall et al. (1983) J. Mol. Appl. Gen. 2: 101), human placental secreted alkaline phosphatase (Cullen and Malim (1992) Methods in Enzymol. 216:362-368).

Accordingly, yet another embodiment of the subject drug screening assays of the present invention provides a recombinant cell, e.g., for carrying out certain of the drug screening methods above, comprising: (i) an expressible recombinant gene encoding a heterologous *HIP* polypeptide whose signal transduction activity is modulated by binding to a *hedgehog* protein; and (ii) a reporter gene construct containing a reporter gene in operative linkage with one or more transcriptional regulatory elements responsive to the signal transduction activity of the *HIP* polypeptide. Still another aspect of the present invention provides a kit for screening test compounds to identify agents which modulate the binding of *hedgehog* proteins with a *hedgehog* receptor, including the above-referenced cell and a preparation of purified *hedgehog* polypeptide.

In still another embodiment of a drug screening, a two hybrid assay (described *supra*) can be generated with a *HIP* polypeptide and target molecule. Drug dependent inhibition or potentiation of the interaction can be scored.

After identifying certain test compounds as potential modulators of one or more bioactivities of a *HIP* protein (such as *hedgehog* binding), the practioner of the subject assay will continue to test the efficacy and specificity of the selected compounds both *in vitro* and *in vivo*. Whether for subsequent *in vivo* testing, or for administration to an animal as an approved drug, agents identified in the subject assay can be formulated in pharmaceutical preparations for *in vivo* administration to an animal, preferably a human.

Another aspect of the present invention relates to a method of inducing and/or maintaining a differentiated state, enhancing survival, and/or inhibiting (or alternatively potentiating) proliferation of a cell, by contacting the cells with an agent which modulates *HIP*-dependent signal transduction pathways. The subject method could be used to generate and/or maintain an array of different tissue both *in vitro* and *in vivo*. A "*HIP* therapeutic," whether inhibitory or potentiating with respect to modulating the activity of a *HIP* protein, can be, as appropriate, any of the preparations described above, including isolated *HIP* polypeptides (including both agonist and antagonist forms), gene therapy constructs, antisense molecules, peptidomimetics, or agents identified in the drug assays provided herein. In certain embodiments, soluble forms of the *HIP* protein including the extracellular ligand-binding domain of the receptor can be provided as a means for antagonizing the binding of a *HIP* ligand to a cell-surface *HIP* receptor. For instance, such forms of the receptor can be used to antagonize the bioactivity of a ligand of the receptor.

The *HIP* therapeutic compounds of the present invention are likely to play an important role in the modulation of cellular proliferation and maintenance of, for example, neuronal, testicular, osteogenic or chondrogenic tissues during disease states. It will also be apparent that, by transient use of modulators of *HIP* activities, *in vivo* reformation of tissue can be accomplished, e.g. in the development and maintenance of organs such as ectodermal patterning, as well as certain mesodermal and endodermal differentiation processes. By controlling the proliferative and differentiative potential for different cells, the subject *HIP* therapeutics can be used to reform injured tissue, or to improve grafting and morphology of transplanted tissue. For instance, *HIP* antagonists and agonists can be employed in a differential manner to regulate different stages of organ repair after physical, chemical or pathological insult. The present method is also applicable to cell culture techniques.

To further illustrate this aspect of the invention, *in vitro* neuronal culture systems have proved to be fundamental and indispensable tools for the study of neural development, as well as the identification of neurotrophic factors such as nerve growth factor (NGF), ciliary trophic factors (CNTF), and brain derived neurotrophic factor (BDNF). Once a neuronal cell has become terminally-differentiated it typically will not change to another terminally differentiated cell-type. However, neuronal cells can nevertheless readily lose their differentiated state. This is commonly observed when they are grown in culture from adult tissue, and when they form a blastema during regeneration. The present method provides a means for ensuring an adequately restrictive environment in order to maintain neuronal cells at various stages of differentiation, and can be employed, for instance, in cell cultures designed to test the specific activities of other trophic factors. In such embodiments of the subject method, the cultured cells can be contacted with a *HIP* therapeutic, e.g., such as an agent identified in the assays described above which potentiate *HIP*-dependent *hedgehog* bioactivities, in order to induce neuronal differentiation (e.g. of a stem cell), or to maintain the integrity of a culture of terminally-differentiated neuronal cells by preventing loss of differentiation. Alternatively, a antagonist of *hedgehog* induction, as certain of the *HIP* homologs of the present invention are expected to be, can be used to prevent differentiation of progenitor cells in culture.

To further illustrate uses of *HIP* therapeutics which may be either *hedgehog* agonists or antagonists, it is noted that intracerebral grafting has emerged as an additional approach to central nervous system therapies. For example, one approach to repairing damaged brain tissues involves the transplantation of cells from fetal or neonatal animals into the adult brain (Dunnett et al. (1987) *J Exp Biol* 123:265-289; and Freund et al. (1985) *J Neurosci* 5:603-616). Fetal neurons from a variety of brain regions can be successfully incorporated into the adult brain, and such grafts can alleviate behavioral defects. For example, movement disorder induced by lesions of dopaminergic projections to the basal ganglia can

be prevented by grafts of embryonic dopaminergic neurons. Complex cognitive functions that are impaired after lesions of the neocortex can also be partially restored by grafts of embryonic cortical cells. The differential use of *hedgehog* agonists and antagonists in the culture can control the timing and type of differentiation accessible by the culture.

In addition to the implantation of cells cultured in the presence of *hedgehog* agonists and antagonists and other *in vitro* uses, yet another aspect of the present invention concerns the therapeutic application of a *HIP* therapeutics to enhance survival of neurons and other neuronal cells in both the central nervous system and the peripheral nervous system. The ability of *hedgehog* protein to regulate neuronal differentiation during development of the nervous system and also presumably in the adult state indicates that certain of the *hedgehog* proteins. and accordingly *HIP* therapeutic which modulate *hedgehog* bioactivities, can be reasonably expected to facilitate control of adult neurons with regard to maintenance, functional performance, and aging of normal cells; repair and regeneration processes in chemically or mechanically lesioned cells; and prevention of degeneration and premature death which result from loss of differentiation in certain pathological conditions. In light of this understanding, the present invention specifically contemplates applications of the subject *HIP* therapeutics to the treatment of (prevention and/or reduction of the severity of) neurological conditions deriving from: (i) acute, subacute, or chronic injury to the nervous system. including traumatic injury, chemical injury, vasal injury and deficits (such as the ischemia resulting from stroke), together with infectious/inflammatory and tumor-induced injury; (ii) aging of the nervous system including Alzheimer's disease; (iii) chronic neurodegenerative diseases of the nervous system, including Parkinson's disease, Huntington's chorea, amylotrophic lateral sclerosis and the like, as well as spinocerebellar degenerations; and (iv) chronic immunological diseases of the nervous system or affecting the nervous system, including multiple sclerosis.

Many neurological disorders are associated with degeneration of discrete populations of neuronal elements and may be treatable with a therapeutic regimen which includes a *HIP* therapeutic that acts as a *hedgehog* agonist. For example, Alzheimer's disease is associated with deficits in several neurotransmitter systems, both those that project to the neocortex and those that reside with the cortex. For instance, the nucleus basalis in patients with Alzheimer's disease have been observed to have a profound (75%) loss of neurons compared to age-matched controls. Although Alzheimer's disease is by far the most common form of dementia, several other disorders can produce dementia. Several of these are degenerative diseases characterized by the death of neurons in various parts of the central nervous system, especially the cerebral cortex. However, some forms of dementia are associated with degeneration of the thalmus or the white matter underlying the cerebral cortex. Here, the cognitive dysfunction results from the isolation of cortical areas by the degeneration of efferents and afferents. Huntington's disease involves the

degeneration of intrastraital and cortical cholinergic neurons and GABAergic neurons. Pick's disease is a severe neuronal degeneration in the neocortex of the frontal and anterior temporal lobes, sometimes accompanied by death of neurons in the striatum. Treatment of patients suffering from such degenerative conditions can include the application of *HIP* therapeutics in order to control, for example, differentiation and apoptotic events which give rise to loss of neurons (e.g. to enhance survival of existing neurons) as well as promote differentiation and repopulation by progenitor cells in the area affected.

In addition to degenerative-induced dementias, a pharmaceutical preparation of one or more of the subject *HIP* therapeutics can be applied opportunely in the treatment of neurodegenerative disorders which have manifestations of tremors and involuntary movements. Parkinson's disease, for example, primarily affects subcortical structures and is characterized by degeneration of the nigrostriatal pathway, raphe nuclei, locus cereleus, and the motor nucleus of vagus. Ballism is typically associated with damage to the subthalmic nucleus, often due to acute vascular accident. Also included are neurogenic and myopathic diseases which ultimately affect the somatic division of the peripheral nervous system and are manifest as neuromuscular disorders. Examples include chronic atrophies such as amyotrophic lateral sclerosis, Guillain-Barre syndrome and chronic peripheral neuropathy, as well as other diseases which can be manifest as progressive bulbar palsies or spinal muscular atrophies. The present method is amenable to the treatment of disorders of the cerebellum which result in hypotonia or ataxia, such as those lesions in the cerebellum which produce disorders in the limbs ipsilateral to the lesion. For instance, a preparation of a *HIP* therapeutic can used to treat a restricted form of cerebellar cortical degeneration involving the anterior lobes (vermis and leg areas) such as is common in alcoholic patients.

In an illustrative embodiment, the subject method is used to treat amyotrophic lateral sclerosis. ALS is a name given to a complex of disorders that comprise upper and lower motor neurons. Patients may present with progressive spinal muscular atrophy, progressive bulbar palsy, primary lateral sclerosis, or a combination of these conditions. The major pathological abnormality is characterized by a selective and progressive degeneration of the lower motor neurons in the spinal cord and the upper motor neurons in the cerebral cortex. The therapeutic application of a *hedgehog* agonist can be used alone, or in conjunction with other neurotrophic factors such as CNTF, BDNF or NGF to prevent and/or reverse motor neuron degeneration in ALS patients.

HIP therapeutics of the present invention can also be used in the treatment of autonomic disorders of the peripheral nervous system, which include disorders affecting the innervation of smooth muscle and endocrine tissue (such as glandular tissue). For instance, the subject method can be used to treat tachycardia or atrial cardiac arrythmias which may

arise from a degenerative condition of the nerves innervating the striated muscle of the heart.

Furthermore, a potential role for certain of the *HIP* therapeutics derives from the role of *hedgehog* proteins in development and maintenance of dendritic processes of axonal neurons. Potential roles for *hedgehog* agonists consequently include guidance for axonal projections and the ability to promote differentiation and/or maintenance of the innervating cells to their axonal processes. Accordingly, compositions comprising *HIP* therapeutics which agonize *hedgehog* activity, may be employed to support the survival and reprojection of several types of ganglionic neurons sympathetic and sensory neurons as well as motor neurons. In particular, such therapeutic compositions may be useful in treatments designed to rescue, for example, various neurons from lesion-induced death as well as guiding reprojection of these neurons after such damage. Such diseases include, but are not limited to, CNS trauma infarction, infection (such as viral infection with varicella-zoster). metabolic disease. nutritional deficiency, toxic agents (such as cisplatin treatment).

Moreover, certain of the *HIP* therapeutics (e.g., which antagonize *hedgehog* induction) may be useful in the selective ablation of sensory neurons, for example, in the treatment of chronic pain syndromes.

As appropriate, *HIP* therapeutics can be used in nerve prostheses for the repair of central and peripheral nerve damage. In particular, where a crushed or severed axon is intubulated by use of a prosthetic device, certain of *HIP* therapeutics can be added to the prosthetic device to increase the rate of growth and regeneration of the dendridic processes. Exemplary nerve guidance channels are described in U.S. patents 5,092,871 and 4,955,892. Accordingly, a severed axonal process can be directed toward the nerve ending from which it was severed by a prosthesis nerve guide.

In another embodiment, the subject method can be used in the treatment of neoplastic or hyperplastic transformations such as may occur in the central nervous system. For instance, certain of the *HIP* therapeutics which induce differentiation of neuronal cells can be utilized to cause such transformed cells to become either post-mitotic or apoptotic. Treatment with a *HIP* therapeutic may facilitate disruption of autocrine loops, such as TGF-β or PDGF autostimulatory loops, which are believed to be involved in the neoplastic transformation of several neuronal tumors. *HIP* therapeutics may, therefore, thus be of use in the treatment of, for example, malignant gliomas, medulloblastomas, neuroectodermal tumors, and ependymonas.

Yet another aspect of the present invention concerns the application of the discovery that *hedgehog* proteins are morphogenic signals involved in other vertebrate organogenic pathways in addition to neuronal differentiation as described above, having apparent roles in other endodermal patterning, as well as both mesodermal and endodermal differentiation

processes. As described in the literature, *Shh* plays a role in proper limb growth and patterning by initiating expression of signaling molecules, including *Bmp-2* in the mesoderm and *Fgf-4* in the ectoderm. Thus, it is contemplated by the invention that compositions comprising caertain of the *HIP* therapeutics can also be utilized for both cell culture and therapeutic methods involving generation and maintenance of non-neuronal tissue.

In one embodiment, the present invention makes use of the discovery that *hedgehog* proteins, such as *Shh*, are apparently involved in controlling the development of stem cells responsible for formation of the digestive tract, liver, lungs, and other organs which derive from the primitive gut. *Shh* serves as an inductive signal from the endoderm to the mesoderm, which is critical to gut morphogenesis. Therefore, for example, *hedgehog* agonists can be employed in the development and maintenance of an artificial liver which can have multiple metabolic functions of a normal liver. In an exemplary embodiment, a *HIP* therapeutic which acts as a *hedgehog* agonist can be used to induce differentiation of digestive tube stem cells to form hepatocyte cultures which can be used to populate extracellular matrices, or which can be encapsulated in biocompatible polymers, to form both implantable and extracorporeal artificial livers.

In another embodiment, therapeutic compositions of *hedgehog* agonists can be utilized in conjunction with transplantation of such artificial livers, as well as embryonic liver structures, to promote intraperitoneal implantation, vascularization, and *in vivo* differentiation and maintenance of the engrafted liver tissue.

In yet another embodiment, *HIP* therapeutics can be employed therapeutically to regulate such organs after physical, chemical or pathological insult. For instance, therapeutic compositions comprising *hedgehog* agonists can be utilized in liver repair subsequent to a partial hepatectomy. Similarly, therapeutic compositions containing *hedgehog* agonists can be used to promote regeneration of lung tissue in the treatment of emphysema.

In still another embodiment of the present invention, compositions comprising *HIP* therapeutics can be used in the *in vitro* generation of skeletal tissue, such as from skeletogenic stem cells, as well as the *in vivo* treatment of skeletal tissue deficiencies. The present invention particularly contemplates the use of *HIP* therapeutics which agonize a *hedgehog* a skeletogenic activity, such as an ability to induce chondrogenesis and/or osteogenesis. By "skeletal tissue deficiency", it is meant a deficiency in bone or other skeletal connective tissue at any site where it is desired to restore the bone or connective tissue, no matter how the deficiency originated, e.g. whether as a result of surgical intervention, removal of tumor, ulceration, implant, fracture, or other traumatic or degenerative conditions.

For instance, the present invention makes available effective therapeutic methods and compositions for restoring cartilage function to a connective tissue. Such methods are useful in. for example, the repair of defects or lesions in cartilage tissue which is the result of degenerative wear such as that which results in arthritis, as well as other mechanical derangements which may be caused by trauma to the tissue, such as a displacement of torn meniscus tissue, meniscectomy, a laxation of a joint by a torn ligament, malignment of joints. bone fracture, or by hereditary disease. The present reparative method is also useful for remodeling cartilage matrix, such as in plastic or reconstructive surgery, as well as periodontal surgery. The present method may also be applied to improving a previous reparative procedure, for example, following surgical repair of a meniscus. ligament, or cartilage. Furthermore, it may prevent the onset or exacerbation of degenerative disease if applied early enough after trauma.

In one embodiment of the present invention, the subject method comprises treating the afflicted connective tissue with a therapeutically sufficient amount of a *hedgehog* agonist, particularly *HIP* therapeutic which agonizes *Ihh* activity, to generate a cartilage repair response in the connective tissue by stimulating the differentiation and/or proliferation of chondrocytes embedded in the tissue. Induction of chondrocytes by treatment with a *hedgehog* agonist can subsequently result in the synthesis of new cartilage matrix by the treated cells. Such connective tissues as articular cartilage, interarticular cartilage (menisci), costal cartilage (connecting the true ribs and the sternum), ligaments, and tendons are particularly amenable to treatment in reconstructive and/or regenerative therapies using the subject method. As used herein, regenerative therapies include treatment of degenerative states which have progressed to the point of which impairment of the tissue is obviously manifest, as well as preventive treatments of tissue where degeneration is in its earliest stages or imminent. The subject method can further be used to prevent the spread of mineralisation into fibrotic tissue by maintaining a constant production of new cartilage.

In an illustrative embodiment, the subject method can be used to treat cartilage of a diarthroidal joint, such as a knee, an ankle, an elbow, a *HIP*, a wrist, a knuckle of either a finger or toe, or a temperomandibular joint. The treatment can be directed to the meniscus of the joint, to the articular cartilage of the joint, or both. To further illustrate, the subject method can be used to treat a degenerative disorder of a knee, such as which might be the result of traumatic injury (e.g., a sports injury or excessive wear) or osteoarthritis. An injection of a *HIP* therapeutic into the joint with, for instance, an arthroscopic needle, can be used to treat the afflicted cartilage. In some instances, the injected agent can be in the form of a hydrogel or other slow release vehicle described above in order to permit a more extended and regular contact of the agent with the treated tissue.

The present invention further contemplates the use of the subject method in the field of cartilage transplantation and prosthetic device therapies. To date, the growth of new cartilage from either transplantation of autologous or allogenic cartilage has been largely unsuccessful. Problems arise, for instance, because the characteristics of cartilage and fibrocartilage varies between different tissue: such as between articular, meniscal cartilage, ligaments, and tendons, between the two ends of the same ligament or tendon, and between the superficial and deep parts of the tissue. The zonal arrangement of these tissues may reflect a gradual change in mechanical properties, and failure occurs when implanted tissue, which has not differentiated under those conditions, lacks the ability to appropriately respond. For instance, when meniscal cartilage is used to repair anterior cruciate ligaments, the tissue undergoes a metaplasia to pure fibrous tissue. By promoting chondrogenesis, the subject method can be used to particularly addresses this problem, by causing the implanted cells to become more adaptive to the new environment and effectively resemble hypertrophic chondrocytes of an earlier developmental stage of the tissue. Thus, the action of chondrogensis in the implanted tissue, as provided by the subject method, and the mechanical forces on the actively remodeling tissue can synergize to produce an improved implant more suitable for the new function to which it is to be put.

In similar fashion, the subject method can be applied to enhancing both the generation of prosthetic cartilage devices and to their implantation. The need for improved treatment has motivated research aimed at creating new cartilage that is based on collagen-glycosaminoglycan templates (Stone et al. (1990) *Clin Orthop Relat Red* 252:129), isolated chondrocytes (Grande et al. (1989) *J Orthop Res* 7:208; and Takigawa et al. (1987) *Bone Miner* 2:449), and chondrocytes attached to natural or synthetic polymers (Walitani et al. (1989) *J Bone Jt Surg* 71B:74; Vacanti et al. (1991) *Plast Reconstr Surg* 88:753; von Schroeder et al. (1991) *J Biomed Mater Res* 25:329; Freed et al. (1993) *J Biomed Mater Res* 27:11; and the Vacanti et al. U.S. Patent No. 5,041,138). For example, chondrocytes can be grown in culture on biodegradable, biocompatible highly porous scaffolds formed from polymers such as polyglycolic acid, polylactic acid, agarose gel, or other polymers which degrade over time as function of hydrolysis of the polymer backbone into innocuous monomers. The matrices are designed to allow adequate nutrient and gas exchange to the cells until engraftment occurs. The cells can be cultured *in vitro* until adequate cell volume and density has developed for the cells to be implanted. One advantage of the matrices is that they can be cast or molded into a desired shape on an individual basis, so that the final product closely resembles the patient's own ear or nose (by way of example), or flexible matrices can be used which allow for manipulation at the time of implantation, as in a joint.

In one embodiment of the subject method, the implants are contacted with a *HIP* therapeutic during the culturing process, such as an *Ihh* agonist, in order to induce and/or maintain differentiated chondrocytes in the culture in order as to further stimulate cartilage

matrix production within the implant. In such a manner, the cultured cells can be caused to maintain a phenotype typical of a chondrogenic cell (i.e. hypertrophic), and hence continue the population of the matrix and production of cartilage tissue.

In another embodiment, the implanted device is treated with a *HIP* therapeutic in order to actively remodel the implanted matrix and to make it more suitable for its intended function. As set out above with respect to tissue transplants, the artificial transplants suffer from the same deficiency of not being derived in a setting which is comparable to the actual mechanical environment in which the matrix is implanted. The activation of the chondrocytes in the matrix by the subject method can allow the implant to acquire characteristics similar to the tissue for which it is intended to replace.

In yet another embodiment, the subject method is used to enhance attachment of prosthetic devices. To illustrate, the subject method can be used in the implantation of a periodontal prosthesis, wherein the treatment of the surrounding connective tissue stimulates formation of periodontal ligament about the prosthesis, as well as inhibits formation of fibrotic tissue proximate the prosthetic device.

In still further embodiments, the subject method can be employed for the generation of bone (osteogenesis) at a site in the animal where such skeletal tissue is deficient. Indian *hedgehog* is particularly associated with the hypertrophic chondrocytes that are ultimately replaced by osteoblasts. For instance, administration of a *HIP* therapeutic of the present invention can be employed as part of a method for treating bone loss in a subject, e.g. to prevent and/or reverse osteoporosis and other osteopenic disorders, as well as to regulate bone growth and maturation. For example, preparations comprising *hedgehog* agonists can be employed, for example, to induce endochondral ossification, at least so far as to facilitate the formation of cartilaginous tissue precursors to form the "model" for ossification. Therapeutic compositions of *HIP* therapeutics can be supplemented, if required, with other osteoinductive factors, such as bone growth factors (e.g. TGF-β factors, such as the bone morphogenetic factors *BMP-2* and *BMP-4*, as well as activin), and may also include, or be administered in combination with, an inhibitor of bone resorption such as estrogen, bisphosphonate, sodium fluoride, calcitonin, or tamoxifen, or related compounds. However, it will be appreciated that *hedgehog* proteins, such as I*hh* and S*hh* are likely to be upstream of BMPs, e.g. treatment with a *hedgehog* agonist will have the advantage of initiating endogenous expression of BMPs along with other factors.

In yet another embodiment, the *HIP* therapeutic of the present invention can be used in the treatment of testicular cells, so as to modulate spermatogenesis. In light of the finding that *hedgehog* proteins are involved in the differentiation and/or proliferation and maintenance of testicular germ cells, *hedgehog* antagonist can be utilized to block the action of a naturally-occurring *hedgehog* protein. In a preferred embodiment, the *HIP*

therapeutic inhibits the biological activity of *Dhh* with respect to spermatogenesis, by competitively binding *hedgehog* in the testis. That is, the *HIP* therapeutic can be administered as a contraceptive formulation. Alternatively, *HIP* therapeutics which agonize the spermatogenic activity of *Dhh* can be used as fertility enhancers. In similar fashion, *hedgehog* agonists and antagonists are potentially useful for modulating normal ovarian function.

Another aspect of the invention features transgenic non-human animals which express a heterologous *HIP* gene of the present invention, and/or which have had one or more genomic *HIP* genes disrupted in at least a tissue or cell-types of the animal. Accordingly, the invention features an animal model for developmental diseases, which animal has one or more *HIP* allele which is mis-expressed. For example, an animal can be generated which has one or more *HIP* alleles deleted or otherwise rendered inactive. Such a model can then be used to study disorders arising from mis-expressed *HIP* genes, as well as for evaluating potential therapies for similar disorders.

The transgenic animals of the present invention all include within a plurality of their cells a transgene of the present invention, which transgene alters the phenotype of the "host cell" with respect to regulation by the *HIP protein*, e.g., of cell growth, death and/or differentiation. Since it is possible to produce transgenic organisms of the invention utilizing one or more of the transgene constructs described herein, a general description will be given of the production of transgenic organisms by referring generally to exogenous genetic material. This general description can be adapted by those skilled in the art in order to incorporate specific transgene sequences into organisms utilizing the methods and materials described herein and those generally known in the art.

In one embodiment, the transgene construct is a knockout construct. Such transgene constructs usually are insertion-type or replacement-type constructs (Hasty et al. (1991) *Mol Cell Biol* 11:4509). The transgene constructs for disruption of a *HIP* gene are designed to facilitate homologous recombination with a portion of the genomic *HIP* gene so as to prevent the functional expression of the endogenous *HIP* gene. In preferred embodiments, the nucleotide sequence used as the knockout construct can be comprised of (1) DNA from some portion of the endogenous *HIP* gene (exon sequence, intron sequence, promoter sequences, etc.) which direct recombination and (2) a marker sequence which is used to detect the presence of the knockout construct in the cell. The knockout construct is inserted into a cell, and integrates with the genomic DNA of the cell in such a position so as to prevent or interrupt transcription of the native *HIP* gene. Such insertion can occur by homologous recombination, i.e., regions of the knockout construct that are homologous to the endogenous *HIP* gene sequence hybridize to the genomic DNA and recombine with the genomic sequences so that the construct is incorporated into the corresponding position of

the genomic DNA. The knockout construct can comprise (1) a full or partial sequence of one or more exons and/or introns of the *HIP* gene to be disrupted, (2) sequences which flank the 5' and 3' ends of the coding sequence of the *HIP* gene, or (3) a combination thereof.

A preferred knockout construct will delete, by targeted homologous recombination, essential structural elements of an endogenous *HIP* gene. For example, the targeting construct can recombine with the genomic *HIP* gene can delete a portion of the coding sequence, and/or essential transcriptional regulatory sequences of the gene.

Alternatively, the knockout construct can be used to interrupt essential structural and/or regulatory elements of an endogenous *HIP* gene by targeted insertion of a polynucleotide sequence. For instance, a knockout construct can recombine with a *HIP* gene and insert a nonhomologous sequence, such as a *neo* expression cassette, into a structural element (e.g., an exon) and/or regulatory element (e.g., enhancer, promoter, intron splice site, polyadenylation site, etc.) to yield a targeted *HIP* allele having an insertional disruption. The inserted nucleic acid can range in size from 1 nucleotide (e.g., to produce a frameshift) to several kilobases or more, and is limited only by the efficiency of the targeting technique.

Depending of the location and characteristics of the disruption, the transgene construct can be used to generate a transgenic animal in which substantially all expression of the targeted *HIP* gene is inhibited in at least a portion of the animal's cells. If only regulatory elements are targeted, some low-level expression of the targeted gene may occur (i.e., the targeted allele is "leaky").

The nucleotide sequence(s) comprising the knockout construct(s) can be obtained using methods well known in the art. Such methods include, for example, screening genomic libraries with *HIP* cDNA probes in order to identify the corresponding genomic *HIP* gene and regulatory sequences. Alternatively, where the cDNA sequence is to be used as part of the knockout construct, the cDNA may be obtained by screening a cDNA library as set out above.

In another embodiment, the "transgenic non-human animals" of the invention are produced by introducing transgenes into the germline of the non-human animal. Embryonal target cells at various developmental stages can be used to introduce transgenes. Different methods are used depending on the stage of development of the embryonal target cell. The specific line(s) of any animal used to practice this invention are selected for general good health, good embryo yields, good pronuclear visibility in the embryo, and good reproductive fitness. In addition, the haplotype is a significant factor. For example, when transgenic mice are to be produced, strains such as C57BL/6 or FVB lines are often used (Jackson Laboratory, Bar Harbor, ME). Preferred strains are those with H-2^b, H-2^d or H-2^q haplotypes such as C57BL/6 or DBA/1. The line(s) used to practice this invention may

themselves be transgenics, and/or may be knockouts (i.e., obtained from animals which have one or more genes partially or completely suppressed) .

In one embodiment, the transgene construct is introduced into a single stage embryo. The zygote is the best target for micro-injection. The use of zygotes as a target for gene transfer has a major advantage in that in most cases the injected DNA will be incorporated into the host gene before the first cleavage (Brinster et al. (1985) *PNAS* 82:4438-4442). As a consequence, all cells of the transgenic animal will carry the incorporated transgene. This will in general also be reflected in the efficient transmission of the transgene to offspring of the founder since 50% of the germ cells will harbor the transgene.

Introduction of the transgene nucleotide sequence into the embryo may be accomplished by any means known in the art such as, for example, microinjection, electroporation, or lipofection. Following introduction of the transgene nucleotide sequence into the embryo, the embryo may be incubated *in vitro* for varying amounts of time, or reimplanted into the surrogate host, or both. In vitro incubation to maturity is within the scope of this invention. One common method in to incubate the embryos *in vitro* for about 1-7 days, depending on the species, and then reimplant them into the surrogate host.

Any technique which allows for the addition of the exogenous genetic material into nucleic genetic material can be utilized so long as it is not destructive to the cell, nuclear membrane or other existing cellular or genetic structures. The exogenous genetic material is preferentially inserted into the nucleic genetic material by microinjection. Microinjection of cells and cellular structures is known and is used in the art.

Reimplantation is accomplished using standard methods. Usually, the surrogate host is anesthetized, and the embryos are inserted into the oviduct. The number of embryos implanted into a particular host will vary by species, but will usually be comparable to the number of off spring the species naturally produces.

Transgenic offspring of the surrogate host may be screened for the presence and/or expression of the transgene by any suitable method. Screening is often accomplished by Southern blot or Northern blot analysis, using a probe that is complementary to at least a portion of the transgene. Western blot analysis using an antibody against the protein encoded by the transgene may be employed as an alternative or additional method for screening for the presence of the transgene product. Typically, DNA is prepared from excised tissue and analyzed by Southern analysis or PCR for the transgene. Alternatively, the tissues or cells believed to express the transgene at the highest levels are tested for the presence and expression of the transgene using Southern analysis or PCR, although any tissues or cell types may be used for this analysis.

Retroviral infection can also be used to introduce transgene into a non-human animal. The developing non-human embryo can be cultured *in vitro* to the blastocyst stage. During this time, the blastomeres can be targets for retroviral infection (Jaenich. R. (1976) *PNAS* 73:1260-1264). Efficient infection of the blastomeres is obtained by enzymatic treatment to remove the zona pellucida (*Manipulating the Mouse Embryo*, Hogan eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1986). The viral vector system used to introduce the transgene is typically a replication-defective retrovirus carrying the transgene (Jahner et al. (1985) *PNAS* 82:6927-6931; Van der Putten et al. (1985) *PNAS* 82:6148-6152). Transfection is easily and efficiently obtained by culturing the blastomeres on a monolayer of virus-producing cells (Van der Putten, *supra*; Stewart et al. (1987) *EMBO J.* 6:383-388). Alternatively, infection can be performed at a later stage. Virus or virus-producing cells can be injected into the blastocoele (Jahner et al. (1982) *Nature* 298:623-628). Most of the founders will be mosaic for the transgene since incorporation occurs only in a subset of the cells which formed the transgenic non-human animal. Further, the founder may contain various retroviral insertions of the transgene at different positions in the genome which generally will segregate in the offspring. In addition, it is also possible to introduce transgenes into the germ line by intrauterine retroviral infection of the midgestation embryo (Jahner et al. (1982) *supra*).

A third type of target cell for transgene introduction is the embryonal stem cell (ES). ES cells are obtained from pre-implantation embryos cultured *in vitro* and fused with embryos (Evans et al. (1981) *Nature* 292:154-156; Bradley et al. (1984) *Nature* 309:255-258; Gossler et al. (1986) *PNAS* 83: 9065-9069; and Robertson et al. (1986) *Nature* 322:445-448). Transgenes can be efficiently introduced into the ES cells by DNA transfection or by retrovirus-mediated transduction. Such transformed ES cells can thereafter be combined with blastocysts from a non-human animal. The ES cells thereafter colonize the embryo and contribute to the germ line of the resulting chimeric animal. For review see Jaenisch, R. (1988) *Science* 240:1468-1474.

In one embodiment, gene targeting, which is a method of using homologous recombination to modify an animal's genome, can be used to introduce changes into cultured embryonic stem cells. By targeting the *HIP* gene in ES cells, these changes can be introduced into the germlines of animals to generate chimeras. The gene targeting procedure is accomplished by introducing into tissue culture cells a DNA targeting construct that includes a segment homologous to a *HIP* locus, and which also includes an intended sequence modification to the *HIP* genomic sequence (e.g., insertion, deletion, point mutation). The treated cells are then screened for accurate targeting to identify and isolate those which have been properly targeted.

Gene targeting in embryonic stem cells is in fact a scheme contemplated by the present invention as a means for disrupting a *HIP* gene function through the use of a targeting transgene construct designed to undergo homologous recombination with *HIP* genomic sequences. Targeting construct can be arranged so that, upon recombination with an element of a *HIP* gene, a positive selection marker is inserted into (or replaces) coding sequences of the targeted *HIP* gene. The inserted sequence functionally disrupts the *HIP* gene, while also providing a positive selection trait.

Generally, the embryonic stem cells (ES cells) used to produce the knockout animals will be of the same species as the knockout animal to be generated. Thus for example, mouse embryonic stem cells will usually be used for generation of a *HIP*-knockout mice.

Embryonic stem cells are generated and maintained using methods well known to the skilled artisan such as those described by Doetschman et al. (1985) *J. Embryol. Exp. Morphol.* 87:27-45). Any line of ES cells can be used, however, the line chosen is typically selected for the ability of the cells to integrate into and become part of the germ line of a developing embryo so as to create germ line transmission of the knockout construct. Thus, any ES cell line that is believed to have this capability is suitable for use herein. The cells are cultured and prepared for knockout construct insertion using methods well known to the skilled artisan, such as those set forth by Robertson in: Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, E.J. Robertson, ed. IRL Press, Washington, D.C. [1987]); by Bradley et al. (1986) *Current Topics in Devel. Biol.* 20:357-371); and by Hogan et al. (Manipulating the Mouse Embryo: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY [1986]) .

Insertion of the knockout construct into the ES cells can be accomplished using a variety of methods well known in the art including for example, electroporation, microinjection, and calcium phosphate treatment. A preferred method of insertion is electroporation .

Each knockout construct to be inserted into the cell must first be in the linear form. Therefore, if the knockout construct has been inserted into a vector, linearization is accomplished by digesting the DNA with a suitable restriction endonuclease selected to cut only within the vector sequence and not within the knockout construct sequence.

For insertion, the knockout construct is added to the ES cells under appropriate conditions for the insertion method chosen, as is known to the skilled artisan. Where more than one construct is to be introduced into the ES cell, each knockout construct can be introduced simultaneously or one at a time.

If the ES cells are to be electroporated, the ES cells and knockout construct DNA are exposed to an electric pulse using an electroporation machine and following the manufacturer's guidelines for use. After electroporation, the ES cells are typically allowed to recover under suitable incubation conditions. The cells are then screened for the presence of the knockout construct .

Screening can be accomplished using a variety of methods. Where the marker gene is an antibiotic resistance gene, the ES cells may be cultured in the presence of an otherwise lethal concentration of antibiotic. Those ES cells that survive have presumably integrated the knockout construct. If the marker gene is other than an antibiotic resistance gene, a Southern blot of the ES cell genomic DNA can be probed with a sequence of DNA designed to hybridize only to the marker sequence Alternatively, PCR can be used. Finally, if the marker gene is a gene that encodes an enzyme whose activity can be detected (e.g., β-galactosidase), the enzyme substrate can be added to the cells under suitable conditions, and the enzymatic activity can be analyzed. One skilled in the art will be familiar with other useful markers and the means for detecting their presence in a given cell. All such markers are contemplated as being included within the scope of the teaching of this invention.

The knockout construct may integrate into several locations in the ES cell genome, and may integrate into a different location in each ES cell's genome due to the occurrence of random insertion events. The desired location of insertion is in a complementary position to the DNA sequence to be knocked out, e.g., the *HIP* coding sequence, transcriptional regulatory sequence, etc. Typically, less than about 1-5 percent of the ES cells that take up the knockout construct will actually integrate the knockout construct in the desired location. To identify those ES cells with proper integration of the knockout construct, total DNA can be extracted from the ES cells using standard methods. The DNA can then be probed on a Southern blot with a probe or probes designed to hybridize in a specific pattern to genomic DNA digested with particular restriction enzyme(s). Alternatively, or additionally, the genomic DNA can be amplified by PCR with probes specifically designed to amplify DNA fragments of a particular size and sequence (i.e., only those cells containing the knockout construct in the proper position will generate DNA fragments of the proper size).

After suitable ES cells containing the knockout construct in the proper location have been identified, the cells can be inserted into an embryo. Insertion may be accomplished in a variety of ways known to the skilled artisan, however a preferred method is by microinjection. For microinjection, about 10-30 cells are collected into a micropipet and injected into embryos that are at the proper stage of development to permit integration of the foreign ES cell containing the knockout construct into the developing embryo. For instance, the transformed ES cells can be microinjected into blastocytes.

After the ES cell has been introduced into the embryo, the embryo may be implanted into the uterus of a pseudopregnant foster mother for gestation. While any foster mother may be used, the foster mother is typically selected for her ability to breed and reproduce well, and for her ability to care for the young. Such foster mothers are typically prepared by mating with vasectomized males of the same species. The stage of the pseudopregnant foster mother is important for successful implantation, and it is species dependent.

Offspring that are born to the foster mother may be screened initially for *HIP* disruptants, DNA from tissue of the offspring may be screened for the presence of the knockout construct using Southern blots and/or PCR as described above. Offspring that appear to be mosaics may then be crossed to each other, if they are believed to carry the knockout construct in their germ line, in order to generate homozygous knockout animals. Homozygotes may be identified by Southern blotting of equivalent amounts of genomic DNA from animals that are the product of this cross, as well as animals that are known heterozygotes and wild type animals.

Other means of identifying and characterizing the knockout offspring are available. For example, Northern blots can be used to probe the mRNA for the presence or absence of transcripts of either the *HIP* gene, the marker gene, or both. In addition, Western blots can be used to assess the (loss of) level of expression of the *HIP* gene knocked out in various tissues of the offspring by probing the Western blot with an antibody against the *HIP* protein, or an antibody against the marker gene product, where this gene is expressed. Finally, *in situ* analysis (such as fixing the cells and labeling with antibody) and/or FACS (fluorescence activated cell sorting) analysis of various cells from the offspring can be conducted using suitable antibodies or *HIP* ligands, e.g., *hedgehog* proteins, to look for the presence or absence of the knockout construct gene product.

Animals containing more than one knockout construct and/or more than one transgene expression construct are prepared in any of several ways. The preferred manner of preparation is to generate a series of animals, each containing a desired transgenic phenotypes. Such animals are bred together through a series of crosses, backcrosses and selections, to ultimately generate a single animal containing all desired knockout constructs and/or expression constructs, where the animal is otherwise congenic (genetically identical) to the wild type except for the presence of the knockout construct(s) and/or transgene(s). Thus, a transgenic avian species can be generated by breeding a first transgenic bird in which the wild-type *HIP* gene is disrupted with a second transgenic bird which has been engineered to express a mutant *HIP* which retains most other biological functions of the receptor.

The transformed animals, their progeny, and cell lines of the present invention provide several important uses that will be readily apparent to one of ordinary skill in the art.

To illustrate, the transgenic animals and cell lines are particularly useful in screening compounds that have potential as prophylactic or therapeutic treatments of diseases such as may involve aberrant expression, or loss, of a *HIP* gene, or aberrant or unwanted activation of receptor signaling. Screening for a useful drug would involve administering the candidate drug over a range of doses to the transgenic animal, and assaying at various time points for the effect(s) of the drug on the disease or disorder being evaluated. Alternatively, or additionally, the drug could be administered prior to or simultaneously with exposure to induction of the disease, if applicable.

In one embodiment, candidate compounds are screened by being administered to the transgenic animal, over a range of doses, and evaluating the animal's physiological response to the compound(s) over time. Administration may be oral, or by suitable injection, depending on the chemical nature of the compound being evaluated. In some cases, it may be appropriate to administer the compound in conjunction with co-factors that would enhance the efficacy of the compound.

In screening cell lines derived from the subject transgenic animals for compounds useful in treating various disorders, the test compound is added to the cell culture medium at the appropriate time, and the cellular response to the compound is evaluated over time using the appropriate biochemical and/or histological assays. In some cases, it may be appropriate to apply the compound of interest to the culture medium in conjunction with co-factors that would enhance the efficacy of the compound.

All of the above-cited references and publications are hereby incorporated by reference.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, numerous equivalents to the specific polypeptides, nucleic acids, methods, assays and reagents described herein. Such equivalents are considered to be within the scope of this invention.

SEQUENCE LISTING

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2103 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..2100

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
ATG CTG AAG ATG CTC TCG TTT AAG CTG CTA CTG CTG GCC GTG GCT CTG        48
Met Leu Lys Met Leu Ser Phe Lys Leu Leu Leu Ala Val Ala Leu
 1               5                  10                  15

GGC TTC TTT GAA GGA GAT GCG AAG TTT GGG GAA AGG AGC GAG GGG AGC        96
Gly Phe Phe Glu Gly Asp Ala Lys Phe Gly Glu Arg Ser Glu Gly Ser
            20                  25                  30

GGA GCG AGA AGG AGA CGG TGC CTG AAT GGG AAC CCC CCA AAG CGC CTA       144
Gly Ala Arg Arg Arg Arg Cys Leu Asn Gly Asn Pro Pro Lys Arg Leu
        35                  40                  45

AAG AGA AGG GAC AGG CGG GTG ATG TCC CAG CTG GAG CTG CTC AGT GGA       192
Lys Arg Arg Asp Arg Arg Val Met Ser Gln Leu Glu Leu Leu Ser Gly
    50                  55                  60

GGA GAG ATC CTG TGT GGT GGC TTC TAC CCA CGA GTA TCT TGC TGC CTG       240
Gly Glu Ile Leu Cys Gly Gly Phe Tyr Pro Arg Val Ser Cys Cys Leu
65                  70                  75                  80

CAG AGT GAC AGC CCT GGA TTG GGG CGT CTG GAG AAC AAG ATC TTT TCT       288
Gln Ser Asp Ser Pro Gly Leu Gly Arg Leu Glu Asn Lys Ile Phe Ser
                85                  90                  95

GCC ACC AAC AAC TCA GAA TGC AGC AGG CTG CTG GAG GAG ATC CAA TGT       336
Ala Thr Asn Asn Ser Glu Cys Ser Arg Leu Leu Glu Glu Ile Gln Cys
            100                 105                 110

GCT CCC TGC TCC CCG CAT TCC CAG AGC CTC TTC TAC ACA CCT GAA AGA       384
Ala Pro Cys Ser Pro His Ser Gln Ser Leu Phe Tyr Thr Pro Glu Arg
        115                 120                 125

GAT GTC CTG GAT GGG GAC CTA GCA CTT CCG CTC CTC TGC AAA GAC TAC       432
Asp Val Leu Asp Gly Asp Leu Ala Leu Pro Leu Leu Cys Lys Asp Tyr
    130                 135                 140

TGC AAA GAA TTC TTT TAT ACT TGC CGA GGC CAT ATT CCA GGT CTT CTT       480
```

```
Cys Lys Glu Phe Phe Tyr Thr Cys Arg Gly His Ile Pro Gly Leu Leu
145                 150                 155                 160

CAA ACA ACT GCT GAT GAA TTT TGC TTT TAC TAT GCA AGA AAA GAT GCT      528
Gln Thr Thr Ala Asp Glu Phe Cys Phe Tyr Tyr Ala Arg Lys Asp Ala
                165                 170                 175

GGG TTA TGC TTT CCA GAC TTC CCG AGA AAG CAA GTC AGA GGA CCA GCA      576
Gly Leu Cys Phe Pro Asp Phe Pro Arg Lys Gln Val Arg Gly Pro Ala
            180                 185                 190

TCT AAC TAC TTG GGC CAG ATG GAA GAC TAC GAG AAA GTG GGG GGG ATC      624
Ser Asn Tyr Leu Gly Gln Met Glu Asp Tyr Glu Lys Val Gly Gly Ile
        195                 200                 205

AGC AGA AAA CAC AAA CAC AAC TGC CTC TGT GTC CAG GAG GTC ATG AGT      672
Ser Arg Lys His Lys His Asn Cys Leu Cys Val Gln Glu Val Met Ser
    210                 215                 220

GGG CTG CGG CAG CCT GTG AGC GCT GTG CAC AGC GGG GAT GGC TCC CAT      720
Gly Leu Arg Gln Pro Val Ser Ala Val His Ser Gly Asp Gly Ser His
225                 230                 235                 240

CGG CTC TTC ATT CTA GAG AAG GAA GGC TAC GTG AAA ATT CTA ACC CCA      768
Arg Leu Phe Ile Leu Glu Lys Glu Gly Tyr Val Lys Ile Leu Thr Pro
                245                 250                 255

GAA GGA GAA CTG TTC AAG GAG CCT TAC TTG GAC ATT CAC AAA CTT GTT      816
Glu Gly Glu Leu Phe Lys Glu Pro Tyr Leu Asp Ile His Lys Leu Val
            260                 265                 270

CAA AGT GGA ATA AAG GGA GGA GAC GAA AGG GGC CTG CTA AGC CTG GCA      864
Gln Ser Gly Ile Lys Gly Gly Asp Glu Arg Gly Leu Leu Ser Leu Ala
        275                 280                 285

TTC CAT CCC AAT TAC AAG AAA AAT GGA AAG CTG TAT GTG TCT TAT ACC      912
Phe His Pro Asn Tyr Lys Lys Asn Gly Lys Leu Tyr Val Ser Tyr Thr
    290                 295                 300

ACC AAC CAG GAA CGG TGG GCT ATT GGG CCT CAC GAC CAC ATT CTT CGG      960
Thr Asn Gln Glu Arg Trp Ala Ile Gly Pro His Asp His Ile Leu Arg
305                 310                 315                 320

GTT GTG GAA TAC ACA GTA TCC AGG AAA AAC CCC CAT CAA GTT GAT GTG     1008
Val Val Glu Tyr Thr Val Ser Arg Lys Asn Pro His Gln Val Asp Val
                325                 330                 335

AGA ACA GCC AGG GTG TTT CTG GAA GTC GCA GAG CTC CAC CGA AAG CAT     1056
Arg Thr Ala Arg Val Phe Leu Glu Val Ala Glu Leu His Arg Lys His
            340                 345                 350

CTT GGG GGA CAG CTG CTC TTT GGT CCT GAT GGC TTT TTG TAC ATC ATC     1104
Leu Gly Gly Gln Leu Leu Phe Gly Pro Asp Gly Phe Leu Tyr Ile Ile
        355                 360                 365

CTT GGG GAT GGT ATG ATC ACA TTG GAT GAC ATG GAA GAG ATG GAT GGG     1152
Leu Gly Asp Gly Met Ile Thr Leu Asp Asp Met Glu Glu Met Asp Gly
    370                 375                 380
```

```
TTA AGT GAC TTC ACA GGC TCT GTG CTG AGG CTG GAC GTG GAC ACC GAC      1200
Leu Ser Asp Phe Thr Gly Ser Val Leu Arg Leu Asp Val Asp Thr Asp
385                 390                 395                 400

ATG TGC AAT GTG CCT TAT TCC ATA CCT CGG AGT AAC CCT CAC TTC AAC      1248
Met Cys Asn Val Pro Tyr Ser Ile Pro Arg Ser Asn Pro His Phe Asn
                405                 410                 415

AGC ACC AAC CAG CCC CCA GAA GTA TTT GCC CAC GGC CTC CAT GAT CCA      1296
Ser Thr Asn Gln Pro Pro Glu Val Phe Ala His Gly Leu His Asp Pro
            420                 425                 430

GGC AGA TGT GCC GTG GAT CGA CAT CCT ACT GAT ATA AAC ATC AAT TTA      1344
Gly Arg Cys Ala Val Asp Arg His Pro Thr Asp Ile Asn Ile Asn Leu
        435                 440                 445

ACA ATA CTT TGC TCA GAT TCC AAC GGG AAA AAC AGG TCA TCA GCC AGA      1392
Thr Ile Leu Cys Ser Asp Ser Asn Gly Lys Asn Arg Ser Ser Ala Arg
    450                 455                 460

ATC CTA CAG ATA ATA AAG GGA AGA GAT TAT GAA AGT GAG CCA TCT CTT      1440
Ile Leu Gln Ile Ile Lys Gly Arg Asp Tyr Glu Ser Glu Pro Ser Leu
465                 470                 475                 480

CTT GAA TTC AAG CCA TTC AGT AAC GGC CCT TTG GTT GGT GGA TTT GTT      1488
Leu Glu Phe Lys Pro Phe Ser Asn Gly Pro Leu Val Gly Gly Phe Val
                485                 490                 495

TAC AGA GGC TGT CAG TCT GAA AGA TTG TAC GGA AGC TAT GTG TTC GGA      1536
Tyr Arg Gly Cys Gln Ser Glu Arg Leu Tyr Gly Ser Tyr Val Phe Gly
            500                 505                 510

GAT CGC AAT GGG AAT TTC TTA ACC CTC CAG CAA AGC CCA GTG ACC AAG      1584
Asp Arg Asn Gly Asn Phe Leu Thr Leu Gln Gln Ser Pro Val Thr Lys
        515                 520                 525

CAA TGG CAA GAA AAG CCG CTC TGC CTG GGT GCC AGC AGC TCC TGT CGA      1632
Gln Trp Gln Glu Lys Pro Leu Cys Leu Gly Ala Ser Ser Ser Cys Arg
    530                 535                 540

GGC TAC TTT TCG GGT CAC ATC TTG GGA TTT GGA GAA GAT GAA TTA GGA      1680
Gly Tyr Phe Ser Gly His Ile Leu Gly Phe Gly Glu Asp Glu Leu Gly
545                 550                 555                 560

GAG GTT TAC ATT CTA TCA AGC AGT AAG AGT ATG ACC CAG ACT CAC AAT      1728
Glu Val Tyr Ile Leu Ser Ser Ser Lys Ser Met Thr Gln Thr His Asn
                565                 570                 575

GGA AAA CTC TAC AAG ATC GTA GAC CCC AAA AGA CCT TTA ATG CCT GAG      1776
Gly Lys Leu Tyr Lys Ile Val Asp Pro Lys Arg Pro Leu Met Pro Glu
            580                 585                 590

GAA TGC AGA GTC ACA GTT CAA CCT GCC CAG CCA CTG ACC TCC GAT TGC      1824
Glu Cys Arg Val Thr Val Gln Pro Ala Gln Pro Leu Thr Ser Asp Cys
        595                 600                 605

TCC CGG CTC TGT CGA AAC GGC TAC TAC ACC CCC ACT GGC AAG TGC TGC      1872
```

```
Ser Arg Leu Cys Arg Asn Gly Tyr Tyr Thr Pro Thr Gly Lys Cys Cys
    610                 615                 620

TGC AGT CCC GGC TGG GAG GGA GAC TTC TGC AGA ATT GCC AAG TGT GAG      1920
Cys Ser Pro Gly Trp Glu Gly Asp Phe Cys Arg Ile Ala Lys Cys Glu
625                 630                 635                 640

CCA GCG TGC CGT CAT GGA GGT GTC TGT GTC AGA CCG AAC AAG TGC CTC      1968
Pro Ala Cys Arg His Gly Gly Val Cys Val Arg Pro Asn Lys Cys Leu
                645                 650                 655

TGT AAA AAG GGC TAT CTT GGT CCT CAA TGT GAA CAA GTG GAC AGG AAC      2016
Cys Lys Lys Gly Tyr Leu Gly Pro Gln Cys Glu Gln Val Asp Arg Asn
            660                 665                 670

GTC CGC AGA GTG ACC AGG GCA GGT ATC CTT GAT CAG ATC ATT GAC ATG      2064
Val Arg Arg Val Thr Arg Ala Gly Ile Leu Asp Gln Ile Ile Asp Met
        675                 680                 685

ACG TCT TAC TTG CTG GAT CTC ACA AGT TAC ATT GTA TAG                  2103
Thr Ser Tyr Leu Leu Asp Leu Thr Ser Tyr Ile Val
    690                 695                 700
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2103 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..2100

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
ATG CTG AAG ATG CTC TCG TTT AAG CTG CTG CTG CTG GCC GTG GCT CTG       48
Met Leu Lys Met Leu Ser Phe Lys Leu Leu Leu Leu Ala Val Ala Leu
 1               5                  10                  15

GGC TTC TTT GAA GGA GAT GCT AAG TTT GGG GAA AGA AAC GAA GGG AGC       96
Gly Phe Phe Glu Gly Asp Ala Lys Phe Gly Glu Arg Asn Glu Gly Ser
            20                  25                  30

GGA GCA AGG AGG AGA AGG TGC CTG AAT GGG AAC CCC CCG AAG CGC CTG      144
Gly Ala Arg Arg Arg Arg Cys Leu Asn Gly Asn Pro Pro Lys Arg Leu
        35                  40                  45

AAA AGG AGA GAC AGG AGG ATG ATG TCC CAG CTG GAG CTG CTG AGT GGG      192
Lys Arg Arg Asp Arg Arg Met Met Ser Gln Leu Glu Leu Leu Ser Gly
    50                  55                  60
```

```
GGA GAG ATG CTG TGC GGT GGC TTC TAC CCT CGG CTG TCC TGC TGC CTG      240
Gly Glu Met Leu Cys Gly Gly Phe Tyr Pro Arg Leu Ser Cys Cys Leu
 65                  70                  75                  80

CGG AGT GAC AGC CCG GGG CTA GGG CGC CTG GAG AAT AAG ATA TTT TCT      288
Arg Ser Asp Ser Pro Gly Leu Gly Arg Leu Glu Asn Lys Ile Phe Ser
                 85                  90                  95

GTT ACC AAC AAC ACA GAA TGT GGG AAG TTA CTG GAG GAA ATC AAA TGT      336
Val Thr Asn Asn Thr Glu Cys Gly Lys Leu Leu Glu Glu Ile Lys Cys
            100                 105                 110

GCA CTT TGC TCT CCA CAT TCT CAA AGC CTG TTC CAC TCA CCT GAG AGA      384
Ala Leu Cys Ser Pro His Ser Gln Ser Leu Phe His Ser Pro Glu Arg
        115                 120                 125

GAA GTC TTG GAA AGA GAC ATA GTA CTT CCT CTG CTC TGC AAA GAC TAT      432
Glu Val Leu Glu Arg Asp Ile Val Leu Pro Leu Leu Cys Lys Asp Tyr
    130                 135                 140

TGC AAA GAA TTC TTT TAC ACT TGC CGA GGC CAT ATT CCA GGT TTC CTT      480
Cys Lys Glu Phe Phe Tyr Thr Cys Arg Gly His Ile Pro Gly Phe Leu
145                 150                 155                 160

CAA ACA ACT GCG GAT GAG TTT TGC TTT TAC TAT GCA AGA AAA GAT GGT      528
Gln Thr Thr Ala Asp Glu Phe Cys Phe Tyr Tyr Ala Arg Lys Asp Gly
                165                 170                 175

GGG TTG TGC TTT CCA GAT TTT CCA AGA AAA CAA GTC AGA GGA CCA GCA      576
Gly Leu Cys Phe Pro Asp Phe Pro Arg Lys Gln Val Arg Gly Pro Ala
            180                 185                 190

TCT AAC TAC TTG GAC CAG ATG GAA GAA TAT GAC AAA GTG GAA GAG ATC      624
Ser Asn Tyr Leu Asp Gln Met Glu Glu Tyr Asp Lys Val Glu Glu Ile
        195                 200                 205

AGC AGA AAG CAC AAA CAC AAC TGC TTC TGT ATT CAG GAG GTT GTG AGT      672
Ser Arg Lys His Lys His Asn Cys Phe Cys Ile Gln Glu Val Val Ser
    210                 215                 220

GGG CTG CGG CAG CCC GTT GGT GCC CTG CAT AGT GGG GAT GGC TCG CAA      720
Gly Leu Arg Gln Pro Val Gly Ala Leu His Ser Gly Asp Gly Ser Gln
225                 230                 235                 240

CGT CTC TTC ATT CTG GAA AAA GAA GGT TAT GTG AAG ATA CTT ACC CCT      768
Arg Leu Phe Ile Leu Glu Lys Glu Gly Tyr Val Lys Ile Leu Thr Pro
                245                 250                 255

GAA GGA GAA ATT TTC AAG GAG CCT TAT TTG GAC ATT CAC AAA CTT GTT      816
Glu Gly Glu Ile Phe Lys Glu Pro Tyr Leu Asp Ile His Lys Leu Val
            260                 265                 270

CAA AGT GGA ATA AAG GGA GGA GAT GAA AGA GGA CTG CTA AGC CTC GCA      864
Gln Ser Gly Ile Lys Gly Gly Asp Glu Arg Gly Leu Leu Ser Leu Ala
        275                 280                 285

TTC CAT CCC AAT TAC AAG AAA AAT GGA AAG TTG TAT GTG TCC TAT ACC      912
Phe His Pro Asn Tyr Lys Lys Asn Gly Lys Leu Tyr Val Ser Tyr Thr
```

```
     290                 295                 300

ACC AAC CAA GAA CGG TGG GCT ATC GGG CCT CAT GAC CAC ATT CTT AGG      960
Thr Asn Gln Glu Arg Trp Ala Ile Gly Pro His Asp His Ile Leu Arg
305                 310                 315                 320

GTT GTG GAA TAC ACA GTA TCC AGA AAA AAT CCA CAC CAA GTT GAT TTG     1008
Val Val Glu Tyr Thr Val Ser Arg Lys Asn Pro His Gln Val Asp Leu
                325                 330                 335

AGA ACA GCC AGA ATC TTT CTT GAA GTT GCA GAA CTC CAC AGA AAG CAT     1056
Arg Thr Ala Arg Ile Phe Leu Glu Val Ala Glu Leu His Arg Lys His
            340                 345                 350

CTG GGA GGA CAA CTG CTC TTT GGC CCT GAC GGC TTT TTG TAC ATC ATT     1104
Leu Gly Gly Gln Leu Leu Phe Gly Pro Asp Gly Phe Leu Tyr Ile Ile
        355                 360                 365

CTT GGT GAT GGG ATG ATT ACA CTG GAT GAT ATG GAA GAA ATG GAT GGG     1152
Leu Gly Asp Gly Met Ile Thr Leu Asp Asp Met Glu Glu Met Asp Gly
    370                 375                 380

TTA AGT GAT TTC ACA GGC TCA GTG CTA CGG CTG GAT GTG GAC ACA GAC     1200
Leu Ser Asp Phe Thr Gly Ser Val Leu Arg Leu Asp Val Asp Thr Asp
385                 390                 395                 400

ATG TGC AAC GTG CCT TAT TCC ATA CCA AGG AGC AAC CCA CAC TTC AAC     1248
Met Cys Asn Val Pro Tyr Ser Ile Pro Arg Ser Asn Pro His Phe Asn
                405                 410                 415

AGC ACC AAC CAG CCC CCC GAA GTG TTT GCT CAT GGG CTC CAC GAT CCA     1296
Ser Thr Asn Gln Pro Pro Glu Val Phe Ala His Gly Leu His Asp Pro
            420                 425                 430

GGC AGA TGT GCT GTG GAT AGA CAT CCC ACT GAT ATA AAC ATC AAT TTA     1344
Gly Arg Cys Ala Val Asp Arg His Pro Thr Asp Ile Asn Ile Asn Leu
        435                 440                 445

ACG ATA CTG TGT TCA GAC TCC AAT GGA AAA AAC AGA TCA TCA GCC AGA     1392
Thr Ile Leu Cys Ser Asp Ser Asn Gly Lys Asn Arg Ser Ser Ala Arg
    450                 455                 460

ATT CTA CAG ATA ATA AAG GGN ARR GAY TAT GAA AGT GAG CCN TCN CTT     1440
Ile Leu Gln Ile Ile Lys Gly Xaa Asp Tyr Glu Ser Glu Pro Ser Leu
465                 470                 475                 480

CTT GAA TTC AAG CCA TTC AGT AAT GGT CCT TTG GTT GGT GGA TTT GTA     1488
Leu Glu Phe Lys Pro Phe Ser Asn Gly Pro Leu Val Gly Gly Phe Val
                485                 490                 495

TAC CGG GGC TGC CAG TCA GAA AGA TTG TAT GGA AGC TAC GTG TTT GGA     1536
Tyr Arg Gly Cys Gln Ser Glu Arg Leu Tyr Gly Ser Tyr Val Phe Gly
            500                 505                 510

GAT CGT AAT GGG AAT TTC CTA ACT CTC CAG CAA AGT CCT GTG ACA AAG     1584
Asp Arg Asn Gly Asn Phe Leu Thr Leu Gln Gln Ser Pro Val Thr Lys
        515                 520                 525
```

```
CAG TGG CAA GAA AAA CCA CTC TGT CTC GGC ACT AGT GGG TCC TGT AGA      1632
Gln Trp Gln Glu Lys Pro Leu Cys Leu Gly Thr Ser Gly Ser Cys Arg
    530                 535                 540

GGC TAC TTT TCC GGT CAC ATC TTG GGA TTT GGA GAA GAT GAA CTA GGT      1680
Gly Tyr Phe Ser Gly His Ile Leu Gly Phe Gly Glu Asp Glu Leu Gly
545                 550                 555                 560

GAA GTT TAC ATT TTA TCA AGC AGT AAA AGT ATG ACC CAG ACT CAC AAT      1728
Glu Val Tyr Ile Leu Ser Ser Ser Lys Ser Met Thr Gln Thr His Asn
                565                 570                 575

GGA AAA CTC TAC AAA ATT GTA GAT CCC AAA AGA CCT TTA ATG CCT GAG      1776
Gly Lys Leu Tyr Lys Ile Val Asp Pro Lys Arg Pro Leu Met Pro Glu
            580                 585                 590

GAA TGC AGA GCC ACG GTA CAA CCT GCA CAG ACA CTG ACT TCA GAG TGC      1824
Glu Cys Arg Ala Thr Val Gln Pro Ala Gln Thr Leu Thr Ser Glu Cys
        595                 600                 605

TCC AGG CTC TGT CGA AAC GGC TAC TGC ACC CCC ACG GGA AAG TGC TGC      1872
Ser Arg Leu Cys Arg Asn Gly Tyr Cys Thr Pro Thr Gly Lys Cys Cys
    610                 615                 620

TGC AGT CCA GGC TGG GAG GGG GAC TTC TGC AGA ACT GCA AAA TGT GAG      1920
Cys Ser Pro Gly Trp Glu Gly Asp Phe Cys Arg Thr Ala Lys Cys Glu
625                 630                 635                 640

CCA GCA TGT CGT CAT GGA GGT GTC TGT GTT AGA CCG AAC AAG TGC CTC      1968
Pro Ala Cys Arg His Gly Gly Val Cys Val Arg Pro Asn Lys Cys Leu
                645                 650                 655

TGT AAA AAA GGA TAT CTT GGT CCT CAA TGT GAA CAA GTG GAC AGA AAC      2016
Cys Lys Lys Gly Tyr Leu Gly Pro Gln Cys Glu Gln Val Asp Arg Asn
            660                 665                 670

ATC CGC AGA GTG ACC AGG GCA GGT ATC CTT GAT CAG ATC ATT GAC ATG      2064
Ile Arg Arg Val Thr Arg Ala Gly Ile Leu Asp Gln Ile Ile Asp Met
        675                 680                 685

ACG TCT TAC TTG CTG GAT CTC ACA AGT TAC ATT GTA TAG                  2103
Thr Ser Tyr Leu Leu Asp Leu Thr Ser Tyr Ile Val
    690                 695                 700
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2085 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS

(B) LOCATION: 1..2082

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
ATG CTC AAG ATG CTG CCG TTC AAG CTG CTG CTG GTG GCC GTG GCT CTG        48
Met Leu Lys Met Leu Pro Phe Lys Leu Leu Leu Val Ala Val Ala Leu
  1               5                  10                  15

TGC TTC TTC GAG GGG GAT GCC AAG TTC GGG GAG AGC GGC GCG CGG AGG        96
Cys Phe Phe Glu Gly Asp Ala Lys Phe Gly Glu Ser Gly Ala Arg Arg
            20                  25                  30

AGA AGG TGC CTC AAC GGG ACC CCC GCG GCG GCT GAA GAA GCG CGA CCG       144
Arg Arg Cys Leu Asn Gly Thr Pro Ala Ala Ala Glu Glu Ala Arg Pro
        35                  40                  45

GCG GCT GCT GTC CCC GGA CCG GGC GGC GCG GAG GCG ATG TGC CGC GGC       192
Ala Ala Ala Val Pro Gly Pro Gly Gly Ala Glu Ala Met Cys Arg Gly
    50                  55                  60

CTC TAC CCG CGC CTC TCC TGC TGC TCC CCG GCC GAC GCG CAG GGG TTG       240
Leu Tyr Pro Arg Leu Ser Cys Cys Ser Pro Ala Asp Ala Gln Gly Leu
65                  70                  75                  80

CTG CAC GCC GGG GCC AAG ATA CTT TCT GTC ACG AAC AAC ACA GAA TGT       288
Leu His Ala Gly Ala Lys Ile Leu Ser Val Thr Asn Asn Thr Glu Cys
                85                  90                  95

GCG AAG CTA CTG GAG GAA ATC AAA TGC GCA CAC TGC TCA CCT CAT GCC       336
Ala Lys Leu Leu Glu Glu Ile Lys Cys Ala His Cys Ser Pro His Ala
            100                 105                 110

CAG AAT CTT TTC CAC TCA CCT GAG AAA GGG GAA ACT TCT GAA AGA GAA       384
Gln Asn Leu Phe His Ser Pro Glu Lys Gly Glu Thr Ser Glu Arg Glu
        115                 120                 125

CTA ACT CTT CCC TAC TTG TGC AAA GAC TAT TGT AAA GAA TTC TAT TAT       432
Leu Thr Leu Pro Tyr Leu Cys Lys Asp Tyr Cys Lys Glu Phe Tyr Tyr
    130                 135                 140

ACT TGC AGA GGT CAC TTA CCA GGT TTT CTC CAA ACT ACA GCT GAT GAG       480
Thr Cys Arg Gly His Leu Pro Gly Phe Leu Gln Thr Thr Ala Asp Glu
145                 150                 155                 160

TTT TGC TTT TAC TAT GCA AGA AAA GAT GGT GGT GTA TGC TTT CCA GAT       528
Phe Cys Phe Tyr Tyr Ala Arg Lys Asp Gly Gly Val Cys Phe Pro Asp
                165                 170                 175

TTT CCA AGA AAA CAA GTG CGA GGG CCA GCT TCT AAC TCC CTG GAC CAC       576
Phe Pro Arg Lys Gln Val Arg Gly Pro Ala Ser Asn Ser Leu Asp His
            180                 185                 190

ATG GAG GAA TAT GAC AAA GAG GAA GAG ATC AGC AGA AAG CAC AAG CAC       624
Met Glu Glu Tyr Asp Lys Glu Glu Glu Ile Ser Arg Lys His Lys His
        195                 200                 205

AAC TGC TTC TGT ATT CAG GAA GTC ATG AGC GGA CTA AGG CAG CCT GTT       672
```

```
Asn Cys Phe Cys Ile Gln Glu Val Met Ser Gly Leu Arg Gln Pro Val
    210                 215                 220

GGA GCG GTA CAT TGT GGG GAT GGA TCT CAT CGC CTC TTT ATT CTT GAG      720
Gly Ala Val His Cys Gly Asp Gly Ser His Arg Leu Phe Ile Leu Glu
225                 230                 235                 240

AAA GAA GGA TAT GTG AAG ATT TTC AGT CCT GAA GGA GAC ATG ATC AAG      768
Lys Glu Gly Tyr Val Lys Ile Phe Ser Pro Glu Gly Asp Met Ile Lys
                245                 250                 255

GAA CCT TTT TTG GAT ATA CAC AAG CTT GTT CAA AGT GGA ATA AAG GGA      816
Glu Pro Phe Leu Asp Ile His Lys Leu Val Gln Ser Gly Ile Lys Gly
            260                 265                 270

GGA GAT GAA AGA GGA CTG TTA AGC CTT GCA TTC CAT CCC AAT TAC AAG      864
Gly Asp Glu Arg Gly Leu Leu Ser Leu Ala Phe His Pro Asn Tyr Lys
        275                 280                 285

AAA AAT GGA AAG CTG TAT GTG TCT TAT ACC ACC AAC CAA GAA CGG TGG      912
Lys Asn Gly Lys Leu Tyr Val Ser Tyr Thr Thr Asn Gln Glu Arg Trp
    290                 295                 300

GCT ATT GGA CCT CAT GAT CAC ATC CTT AGG GTG GTA GAA TAC ACA GTA      960
Ala Ile Gly Pro His Asp His Ile Leu Arg Val Val Glu Tyr Thr Val
305                 310                 315                 320

TCC AGG AAA AAT CCA CAA CAA GTT GAT ATA AGA ACA GCC AGA GTG TTT     1008
Ser Arg Lys Asn Pro Gln Gln Val Asp Ile Arg Thr Ala Arg Val Phe
                325                 330                 335

TTA GAA GTA GCA GAA CTA CAT CGA AAA CAT CTA GGA GGG CAG CTT CTG     1056
Leu Glu Val Ala Glu Leu His Arg Lys His Leu Gly Gly Gln Leu Leu
            340                 345                 350

TTT GGC CCA GAT GGT TTC TTA TAC GTT TTC CTT GGA GAT GGC ATG ATT     1104
Phe Gly Pro Asp Gly Phe Leu Tyr Val Phe Leu Gly Asp Gly Met Ile
        355                 360                 365

ACC CTC GAC GAT ATG GAA GAA ATG GAT GGT TTA AGC GAT TTT ACA GGT     1152
Thr Leu Asp Asp Met Glu Glu Met Asp Gly Leu Ser Asp Phe Thr Gly
    370                 375                 380

TCT GTA TTA CGC CTC GAT GTA AAT ACT GAC CTG TGC AGT GTC CCT TAT     1200
Ser Val Leu Arg Leu Asp Val Asn Thr Asp Leu Cys Ser Val Pro Tyr
385                 390                 395                 400

TCC ATA CCA CGG AGC AAC CCA CAT TTT AAT AGC ACA AAC CAA CCT CCT     1248
Ser Ile Pro Arg Ser Asn Pro His Phe Asn Ser Thr Asn Gln Pro Pro
                405                 410                 415

GAA ATT TTT GCA CAC GGA CTC CAC AAT CCA GGC CGA TGT GCT GTG GAT     1296
Glu Ile Phe Ala His Gly Leu His Asn Pro Gly Arg Cys Ala Val Asp
            420                 425                 430

CAC CAC CCA GCA GAT GTA AAC ATC AAT TTA ACA ATA CTT TGC TCA GAT     1344
His His Pro Ala Asp Val Asn Ile Asn Leu Thr Ile Leu Cys Ser Asp
        435                 440                 445
```

```
TCA AAT GGA AAG AAC AGA TCT TCA GCA AGA ATC TTA CAG ATA ATA AAG      1392
Ser Asn Gly Lys Asn Arg Ser Ser Ala Arg Ile Leu Gln Ile Ile Lys
    450                 455                 460

GGT AAA GAC TAT GAA AGT GAG CCT TCA CTT TTA GAA TTC AAA CCA TTC      1440
Gly Lys Asp Tyr Glu Ser Glu Pro Ser Leu Leu Glu Phe Lys Pro Phe
465                 470                 475                 480

AGC AGT GGA GCG TTG GTC GGT GGA TTT GTC TAT CGA GGT TGC CAG TCT      1488
Ser Ser Gly Ala Leu Val Gly Gly Phe Val Tyr Arg Gly Cys Gln Ser
                485                 490                 495

GAA AGG CTC TAC GGA AGT TAT GTA TTT GGA GAC CGC AAT GGA AAT TTT      1536
Glu Arg Leu Tyr Gly Ser Tyr Val Phe Gly Asp Arg Asn Gly Asn Phe
            500                 505                 510

TTA ACG CTG CAA CAG AAT CCT GCA ACT AAA CAG TGG CAA GAG AAA CCC      1584
Leu Thr Leu Gln Gln Asn Pro Ala Thr Lys Gln Trp Gln Glu Lys Pro
        515                 520                 525

CTC TGT CTT GGC AAC AGC GGT TCA TGT AGA GGT TTC TTT TCA GGC CCT      1632
Leu Cys Leu Gly Asn Ser Gly Ser Cys Arg Gly Phe Phe Ser Gly Pro
    530                 535                 540

GTC TTG GGA TTT GGT GAA GAT GAA CTA GGC GAG ATT TAC ATA TTA TCA      1680
Val Leu Gly Phe Gly Glu Asp Glu Leu Gly Glu Ile Tyr Ile Leu Ser
545                 550                 555                 560

AGC AGT AAA AGT ATG ACA CAG ACT CAC AAT GGA AAA CTC TAC AAG ATC      1728
Ser Ser Lys Ser Met Thr Gln Thr His Asn Gly Lys Leu Tyr Lys Ile
                565                 570                 575

ATT GAC CCA AAA AGG CCT TTA GTT CCT GAA GAA TGC AAA AGA ACA GCT      1776
Ile Asp Pro Lys Arg Pro Leu Val Pro Glu Glu Cys Lys Arg Thr Ala
            580                 585                 590

CGG TCG GCA CAG ATA CTG ACA TCT GAA TGC TCA AGG CAC TGC CGG AAT      1824
Arg Ser Ala Gln Ile Leu Thr Ser Glu Cys Ser Arg His Cys Arg Asn
        595                 600                 605

GGG CAC TGC ACA CCC ACA GGA AAA TGC TGC TGT AAT CAA GGC TGG GAA      1872
Gly His Cys Thr Pro Thr Gly Lys Cys Cys Cys Asn Gln Gly Trp Glu
    610                 615                 620

GGA GAG TTC TGC AGA ACT GCA AAG TGT GAC CCA GCA TGT CGA CAT GGA      1920
Gly Glu Phe Cys Arg Thr Ala Lys Cys Asp Pro Ala Cys Arg His Gly
625                 630                 635                 640

GGT GTC TGT GTA AGG CCT AAT AAA TGC TTA TGT AAA AAA GGC TAT CTT      1968
Gly Val Cys Val Arg Pro Asn Lys Cys Leu Cys Lys Lys Gly Tyr Leu
                645                 650                 655

GGC CCC CAG TGT GAA CAA TTG GAT TTA AAC TTC CGA AAA GTT ACA AGG      2016
Gly Pro Gln Cys Glu Gln Leu Asp Leu Asn Phe Arg Lys Val Thr Arg
            660                 665                 670

CCA GGT ATT CTT GAT CAG ATC CTA AAC ATG ACA TCC TAC TTG CTG GAT      2064
```

```
Pro Gly Ile Leu Asp Gln Ile Leu Asn Met Thr Ser Tyr Leu Leu Asp
        675                 680                 685

CTA ACC AGC TAT ATT GTA TAG                                      2085
Leu Thr Ser Tyr Ile Val
    690
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 173 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..171

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
CAG GAG ATC CAT AGT GGT CTT CAA CAA CCT GTT GGC GTG GTG CAT TGT      48
Gln Glu Ile His Ser Gly Leu Gln Gln Pro Val Gly Val Val His Cys
  1               5                  10                  15

GGA GAT GGA TCG CAG CGG CTT TTT ATA TTG GAG AGG GAA GGC TTT GTG      96
Gly Asp Gly Ser Gln Arg Leu Phe Ile Leu Glu Arg Glu Gly Phe Val
            20                  25                  30

TGG ATC CTC ACA CAT GAC ATG GAA CTC CTA AAA GAG CCT TTT CTG GAC     144
Trp Ile Leu Thr His Asp Met Glu Leu Leu Lys Glu Pro Phe Leu Asp
        35                  40                  45

ATT CAT AAG CTG GTA CAA AGT GGT TTA AA                              173
Ile His Lys Leu Val Gln Ser Gly Leu
    50                  55
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 700 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
Met Leu Lys Met Leu Ser Phe Lys Leu Leu Leu Leu Ala Val Ala Leu
  1               5                  10                  15

Gly Phe Phe Glu Gly Asp Ala Lys Phe Gly Glu Arg Ser Glu Gly Ser
```

```
                20                  25                  30

Gly Ala Arg Arg Arg Arg Cys Leu Asn Gly Asn Pro Pro Lys Arg Leu
        35                  40                  45

Lys Arg Arg Asp Arg Arg Val Met Ser Gln Leu Glu Leu Leu Ser Gly
    50                  55                  60

Gly Glu Ile Leu Cys Gly Gly Phe Tyr Pro Arg Val Ser Cys Cys Leu
65                  70                  75                  80

Gln Ser Asp Ser Pro Gly Leu Gly Arg Leu Glu Asn Lys Ile Phe Ser
                85                  90                  95

Ala Thr Asn Asn Ser Glu Cys Ser Arg Leu Leu Glu Glu Ile Gln Cys
            100                 105                 110

Ala Pro Cys Ser Pro His Ser Gln Ser Leu Phe Tyr Thr Pro Glu Arg
        115                 120                 125

Asp Val Leu Asp Gly Asp Leu Ala Leu Pro Leu Leu Cys Lys Asp Tyr
    130                 135                 140

Cys Lys Glu Phe Phe Tyr Thr Cys Arg Gly His Ile Pro Gly Leu Leu
145                 150                 155                 160

Gln Thr Thr Ala Asp Glu Phe Cys Phe Tyr Tyr Ala Arg Lys Asp Ala
                165                 170                 175

Gly Leu Cys Phe Pro Asp Phe Pro Arg Lys Gln Val Arg Gly Pro Ala
            180                 185                 190

Ser Asn Tyr Leu Gly Gln Met Glu Asp Tyr Glu Lys Val Gly Gly Ile
        195                 200                 205

Ser Arg Lys His Lys His Asn Cys Leu Cys Val Gln Glu Val Met Ser
    210                 215                 220

Gly Leu Arg Gln Pro Val Ser Ala Val His Ser Gly Asp Gly Ser His
225                 230                 235                 240

Arg Leu Phe Ile Leu Glu Lys Glu Gly Tyr Val Lys Ile Leu Thr Pro
                245                 250                 255

Glu Gly Glu Leu Phe Lys Glu Pro Tyr Leu Asp Ile His Lys Leu Val
            260                 265                 270

Gln Ser Gly Ile Lys Gly Gly Asp Glu Arg Gly Leu Leu Ser Leu Ala
        275                 280                 285

Phe His Pro Asn Tyr Lys Lys Asn Gly Lys Leu Tyr Val Ser Tyr Thr
    290                 295                 300

Thr Asn Gln Glu Arg Trp Ala Ile Gly Pro His Asp His Ile Leu Arg
305                 310                 315                 320

Val Val Glu Tyr Thr Val Ser Arg Lys Asn Pro His Gln Val Asp Val
                325                 330                 335
```

Arg Thr Ala Arg Val Phe Leu Glu Val Ala Glu Leu His Arg Lys His
340 345 350

Leu Gly Gly Gln Leu Leu Phe Gly Pro Asp Gly Phe Leu Tyr Ile Ile
355 360 365

Leu Gly Asp Gly Met Ile Thr Leu Asp Asp Met Glu Glu Met Asp Gly
370 375 380

Leu Ser Asp Phe Thr Gly Ser Val Leu Arg Leu Asp Val Asp Thr Asp
385 390 395 400

Met Cys Asn Val Pro Tyr Ser Ile Pro Arg Ser Asn Pro His Phe Asn
405 410 415

Ser Thr Asn Gln Pro Pro Glu Val Phe Ala His Gly Leu His Asp Pro
420 425 430

Gly Arg Cys Ala Val Asp Arg His Pro Thr Asp Ile Asn Ile Asn Leu
435 440 445

Thr Ile Leu Cys Ser Asp Ser Asn Gly Lys Asn Arg Ser Ser Ala Arg
450 455 460

Ile Leu Gln Ile Ile Lys Gly Arg Asp Tyr Glu Ser Glu Pro Ser Leu
465 470 475 480

Leu Glu Phe Lys Pro Phe Ser Asn Gly Pro Leu Val Gly Gly Phe Val
485 490 495

Tyr Arg Gly Cys Gln Ser Glu Arg Leu Tyr Gly Ser Tyr Val Phe Gly
500 505 510

Asp Arg Asn Gly Asn Phe Leu Thr Leu Gln Gln Ser Pro Val Thr Lys
515 520 525

Gln Trp Gln Glu Lys Pro Leu Cys Leu Gly Ala Ser Ser Ser Cys Arg
530 535 540

Gly Tyr Phe Ser Gly His Ile Leu Gly Phe Gly Glu Asp Glu Leu Gly
545 550 555 560

Glu Val Tyr Ile Leu Ser Ser Ser Lys Ser Met Thr Gln Thr His Asn
565 570 575

Gly Lys Leu Tyr Lys Ile Val Asp Pro Lys Arg Pro Leu Met Pro Glu
580 585 590

Glu Cys Arg Val Thr Val Gln Pro Ala Gln Pro Leu Thr Ser Asp Cys
595 600 605

Ser Arg Leu Cys Arg Asn Gly Tyr Tyr Thr Pro Thr Gly Lys Cys Cys
610 615 620

Cys Ser Pro Gly Trp Glu Gly Asp Phe Cys Arg Ile Ala Lys Cys Glu
625 630 635 640

```
Pro Ala Cys Arg His Gly Gly Val Cys Val Arg Pro Asn Lys Cys Leu
                645                 650                 655

Cys Lys Lys Gly Tyr Leu Gly Pro Gln Cys Glu Gln Val Asp Arg Asn
            660                 665                 670

Val Arg Arg Val Thr Arg Ala Gly Ile Leu Asp Gln Ile Ile Asp Met
        675                 680                 685

Thr Ser Tyr Leu Leu Asp Leu Thr Ser Tyr Ile Val
    690                 695                 700
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 700 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Met Leu Lys Met Leu Ser Phe Lys Leu Leu Leu Leu Ala Val Ala Leu
  1               5                  10                  15

Gly Phe Phe Glu Gly Asp Ala Lys Phe Gly Glu Arg Asn Glu Gly Ser
            20                  25                  30

Gly Ala Arg Arg Arg Arg Cys Leu Asn Gly Asn Pro Pro Lys Arg Leu
        35                  40                  45

Lys Arg Arg Asp Arg Arg Met Met Ser Gln Leu Glu Leu Leu Ser Gly
    50                  55                  60

Gly Glu Met Leu Cys Gly Gly Phe Tyr Pro Arg Leu Ser Cys Cys Leu
65                  70                  75                  80

Arg Ser Asp Ser Pro Gly Leu Gly Arg Leu Glu Asn Lys Ile Phe Ser
                85                  90                  95

Val Thr Asn Asn Thr Glu Cys Gly Lys Leu Leu Glu Glu Ile Lys Cys
            100                 105                 110

Ala Leu Cys Ser Pro His Ser Gln Ser Leu Phe His Ser Pro Glu Arg
        115                 120                 125

Glu Val Leu Glu Arg Asp Ile Val Leu Pro Leu Leu Cys Lys Asp Tyr
    130                 135                 140

Cys Lys Glu Phe Phe Tyr Thr Cys Arg Gly His Ile Pro Gly Phe Leu
145                 150                 155                 160

Gln Thr Thr Ala Asp Glu Phe Cys Phe Tyr Tyr Ala Arg Lys Asp Gly
                165                 170                 175
```

Gly Leu Cys Phe Pro Asp Phe Pro Arg Lys Gln Val Arg Gly Pro Ala
180 185 190

Ser Asn Tyr Leu Asp Gln Met Glu Glu Tyr Asp Lys Val Glu Glu Ile
195 200 205

Ser Arg Lys His Lys His Asn Cys Phe Cys Ile Gln Glu Val Val Ser
210 215 220

Gly Leu Arg Gln Pro Val Gly Ala Leu His Ser Gly Asp Gly Ser Gln
225 230 235 240

Arg Leu Phe Ile Leu Glu Lys Glu Gly Tyr Val Lys Ile Leu Thr Pro
245 250 255

Glu Gly Glu Ile Phe Lys Glu Pro Tyr Leu Asp Ile His Lys Leu Val
260 265 270

Gln Ser Gly Ile Lys Gly Gly Asp Glu Arg Gly Leu Leu Ser Leu Ala
275 280 285

Phe His Pro Asn Tyr Lys Lys Asn Gly Lys Leu Tyr Val Ser Tyr Thr
290 295 300

Thr Asn Gln Glu Arg Trp Ala Ile Gly Pro His Asp His Ile Leu Arg
305 310 315 320

Val Val Glu Tyr Thr Val Ser Arg Lys Asn Pro His Gln Val Asp Leu
325 330 335

Arg Thr Ala Arg Ile Phe Leu Glu Val Ala Glu Leu His Arg Lys His
340 345 350

Leu Gly Gly Gln Leu Leu Phe Gly Pro Asp Gly Phe Leu Tyr Ile Ile
355 360 365

Leu Gly Asp Gly Met Ile Thr Leu Asp Asp Met Glu Glu Met Asp Gly
370 375 380

Leu Ser Asp Phe Thr Gly Ser Val Leu Arg Leu Asp Val Asp Thr Asp
385 390 395 400

Met Cys Asn Val Pro Tyr Ser Ile Pro Arg Ser Asn Pro His Phe Asn
405 410 415

Ser Thr Asn Gln Pro Pro Glu Val Phe Ala His Gly Leu His Asp Pro
420 425 430

Gly Arg Cys Ala Val Asp Arg His Pro Thr Asp Ile Asn Ile Asn Leu
435 440 445

Thr Ile Leu Cys Ser Asp Ser Asn Gly Lys Asn Arg Ser Ser Ala Arg
450 455 460

Ile Leu Gln Ile Ile Lys Gly Xaa Asp Tyr Glu Ser Glu Pro Ser Leu
465 470 475 480

Leu Glu Phe Lys Pro Phe Ser Asn Gly Pro Leu Val Gly Gly Phe Val

485 490 495

Tyr Arg Gly Cys Gln Ser Glu Arg Leu Tyr Gly Ser Tyr Val Phe Gly
500 505 510

Asp Arg Asn Gly Asn Phe Leu Thr Leu Gln Gln Ser Pro Val Thr Lys
515 520 525

Gln Trp Gln Glu Lys Pro Leu Cys Leu Gly Thr Ser Gly Ser Cys Arg
530 535 540

Gly Tyr Phe Ser Gly His Ile Leu Gly Phe Gly Glu Asp Glu Leu Gly
545 550 555 560

Glu Val Tyr Ile Leu Ser Ser Ser Lys Ser Met Thr Gln Thr His Asn
565 570 575

Gly Lys Leu Tyr Lys Ile Val Asp Pro Lys Arg Pro Leu Met Pro Glu
580 585 590

Glu Cys Arg Ala Thr Val Gln Pro Ala Gln Thr Leu Thr Ser Glu Cys
595 600 605

Ser Arg Leu Cys Arg Asn Gly Tyr Cys Thr Pro Thr Gly Lys Cys Cys
610 615 620

Cys Ser Pro Gly Trp Glu Gly Asp Phe Cys Arg Thr Ala Lys Cys Glu
625 630 635 640

Pro Ala Cys Arg His Gly Gly Val Cys Val Arg Pro Asn Lys Cys Leu
645 650 655

Cys Lys Lys Gly Tyr Leu Gly Pro Gln Cys Glu Gln Val Asp Arg Asn
660 665 670

Ile Arg Arg Val Thr Arg Ala Gly Ile Leu Asp Gln Ile Ile Asp Met
675 680 685

Thr Ser Tyr Leu Leu Asp Leu Thr Ser Tyr Ile Val
690 695 700

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 694 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Leu Lys Met Leu Pro Phe Lys Leu Leu Leu Val Ala Val Ala Leu
1 5 10 15

Cys Phe Phe Glu Gly Asp Ala Lys Phe Gly Glu Ser Gly Ala Arg Arg
20 25 30

```
Arg Arg Cys Leu Asn Gly Thr Pro Ala Ala Ala Glu Glu Ala Arg Pro
          35                  40                  45

Ala Ala Ala Val Pro Gly Pro Gly Gly Ala Glu Ala Met Cys Arg Gly
      50                  55                  60

Leu Tyr Pro Arg Leu Ser Cys Cys Ser Pro Ala Asp Ala Gln Gly Leu
 65                  70                  75                  80

Leu His Ala Gly Ala Lys Ile Leu Ser Val Thr Asn Asn Thr Glu Cys
                 85                  90                  95

Ala Lys Leu Leu Glu Glu Ile Lys Cys Ala His Cys Ser Pro His Ala
            100                 105                 110

Gln Asn Leu Phe His Ser Pro Glu Lys Gly Glu Thr Ser Glu Arg Glu
        115                 120                 125

Leu Thr Leu Pro Tyr Leu Cys Lys Asp Tyr Cys Lys Glu Phe Tyr Tyr
    130                 135                 140

Thr Cys Arg Gly His Leu Pro Gly Phe Leu Gln Thr Thr Ala Asp Glu
145                 150                 155                 160

Phe Cys Phe Tyr Tyr Ala Arg Lys Asp Gly Gly Val Cys Phe Pro Asp
                165                 170                 175

Phe Pro Arg Lys Gln Val Arg Gly Pro Ala Ser Asn Ser Leu Asp His
            180                 185                 190

Met Glu Glu Tyr Asp Lys Glu Glu Glu Ile Ser Arg Lys His Lys His
        195                 200                 205

Asn Cys Phe Cys Ile Gln Glu Val Met Ser Gly Leu Arg Gln Pro Val
    210                 215                 220

Gly Ala Val His Cys Gly Asp Gly Ser His Arg Leu Phe Ile Leu Glu
225                 230                 235                 240

Lys Glu Gly Tyr Val Lys Ile Phe Ser Pro Glu Gly Asp Met Ile Lys
                245                 250                 255

Glu Pro Phe Leu Asp Ile His Lys Leu Val Gln Ser Gly Ile Lys Gly
            260                 265                 270

Gly Asp Glu Arg Gly Leu Leu Ser Leu Ala Phe His Pro Asn Tyr Lys
        275                 280                 285

Lys Asn Gly Lys Leu Tyr Val Ser Tyr Thr Thr Asn Gln Glu Arg Trp
    290                 295                 300

Ala Ile Gly Pro His Asp His Ile Leu Arg Val Val Glu Tyr Thr Val
305                 310                 315                 320

Ser Arg Lys Asn Pro Gln Gln Val Asp Ile Arg Thr Ala Arg Val Phe
                325                 330                 335
```

Leu Glu Val Ala Glu Leu His Arg Lys His Leu Gly Gly Gln Leu Leu
340 345 350

Phe Gly Pro Asp Gly Phe Leu Tyr Val Phe Leu Gly Asp Gly Met Ile
355 360 365

Thr Leu Asp Asp Met Glu Glu Met Asp Gly Leu Ser Asp Phe Thr Gly
370 375 380

Ser Val Leu Arg Leu Asp Val Asn Thr Asp Leu Cys Ser Val Pro Tyr
385 390 395 400

Ser Ile Pro Arg Ser Asn Pro His Phe Asn Ser Thr Asn Gln Pro Pro
405 410 415

Glu Ile Phe Ala His Gly Leu His Asn Pro Gly Arg Cys Ala Val Asp
420 425 430

His His Pro Ala Asp Val Asn Ile Asn Leu Thr Ile Leu Cys Ser Asp
435 440 445

Ser Asn Gly Lys Asn Arg Ser Ser Ala Arg Ile Leu Gln Ile Ile Lys
450 455 460

Gly Lys Asp Tyr Glu Ser Glu Pro Ser Leu Leu Glu Phe Lys Pro Phe
465 470 475 480

Ser Ser Gly Ala Leu Val Gly Gly Phe Val Tyr Arg Gly Cys Gln Ser
485 490 495

Glu Arg Leu Tyr Gly Ser Tyr Val Phe Gly Asp Arg Asn Gly Asn Phe
500 505 510

Leu Thr Leu Gln Gln Asn Pro Ala Thr Lys Gln Trp Gln Glu Lys Pro
515 520 525

Leu Cys Leu Gly Asn Ser Gly Ser Cys Arg Gly Phe Phe Ser Gly Pro
530 535 540

Val Leu Gly Phe Gly Glu Asp Glu Leu Gly Glu Ile Tyr Ile Leu Ser
545 550 555 560

Ser Ser Lys Ser Met Thr Gln Thr His Asn Gly Lys Leu Tyr Lys Ile
565 570 575

Ile Asp Pro Lys Arg Pro Leu Val Pro Glu Glu Cys Lys Arg Thr Ala
580 585 590

Arg Ser Ala Gln Ile Leu Thr Ser Glu Cys Ser Arg His Cys Arg Asn
595 600 605

Gly His Cys Thr Pro Thr Gly Lys Cys Cys Cys Asn Gln Gly Trp Glu
610 615 620

Gly Glu Phe Cys Arg Thr Ala Lys Cys Asp Pro Ala Cys Arg His Gly
625 630 635 640

Gly Val Cys Val Arg Pro Asn Lys Cys Leu Cys Lys Lys Gly Tyr Leu

645 650 655

Gly Pro Gln Cys Glu Gln Leu Asp Leu Asn Phe Arg Lys Val Thr Arg
660 665 670

Pro Gly Ile Leu Asp Gln Ile Leu Asn Met Thr Ser Tyr Leu Leu Asp
675 680 685

Leu Thr Ser Tyr Ile Val
690

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 57 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Gln Glu Ile His Ser Gly Leu Gln Gln Pro Val Gly Val Val His Cys
1 5 10 15

Gly Asp Gly Ser Gln Arg Leu Phe Ile Leu Glu Arg Glu Gly Phe Val
20 25 30

Trp Ile Leu Thr His Asp Met Glu Leu Leu Lys Glu Pro Phe Leu Asp
35 40 45

Ile His Lys Leu Val Gln Ser Gly Leu
50 55

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 444 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ATGCTGAAGA TGCTCTCGTT TAAGCTGCTG CTGCTGGCCG TGGCTCTGGG CTTCTTTGAA 60

GGAGATGCTA AGTTTGGGGA AAGAAACGAA GGGAGCGGAG CAAGGAGGAG AAGGTGCCTG 120

AATGGGAACC CCCCGAAGCG CCTGAAAAGG AGAGACAGGA GGATGATGTC CCAGCTGGAG 180

CTGCTGAGTG GGGGAGAGAT GCTGTGCGGT GGCTTCTACC CTCGGCTGTC CTGCTGCCTG 240

CGGAGTGACA GCCCGGGGCT AGGGCGCCTG GAGAATAAGA TATTTTCTGT TACCAACAAC 300

ACAGAATGTG GGAAGTTACT GGAGGAAATC AAATGTGCAC TTTGCTCTCC ACATTCTCAA 360

AGCCTGTTCC ACTCACCTGA GAGAGAAGTC TTGGAAAGAG ACATAGTACT TCCTCTGCTC 420

TGCAAAGACT ATTGCAAAGA ATTC 444

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 958 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GAATTCTTTT ACACTTGCCG AGGCCATATT CCAGGTTTCC TTCAAACAAC TGCGGATGAG 60

TTTTGCTTTT ACTATGCAAG AAAAGATGGT GGGTTGTGCT TTCCAGATTT TCCAAGAAAA 120

CAAGTCAGAG GACCAGCATC TAACTACTTG GACCAGATGG AAGAATATGA CAAAGTGGAA 180

GAGATCAGCA GAAAGCACAA ACACAACTGC TTCTGTATTC AGGAGGTTGT GAGTGGGCTG 240

CGGCAGCCCG TTGGTGCCCT GCATAGTGGG GATGGCTCGC AACGTCTCTT CATTCTGGAA 300

AAAGAAGGTT ATGTGAAGAT ACTTACCCCT GAAGGAGAAA TTTTCAAGGA GCCTTATTTG 360

GACATTCACA AACTTGTTCA AAGTGGAATA AAGGGAGGAG ATGAAAGAGG ACTGCTAAGC 420

CTCGCATTCC ATCCCAATTA CAAGAAAAAT GGAAAGTTGT ATGTGTCCTA TACCACCAAC 480

CAAGAACGGT GGGCTATCGG GCCTCATGAC CACATTCTTA GGGTTGTGGA ATACACAGTA 540

TCCAGAAAAA ATCCACACCA AGTTGATTTG AGAACAGCCA GAATCTTTCT TGAAGTTGCA 600

GAACTCCACA GAAAGCATCT GGGAGGACAA CTGCTCTTTG GCCCTGACGG CTTTTTGTAC 660

ATCATTCTTG GTGATGGGAT GATTACACTG GATGATATGG AAGAAATGGA TGGGTTAAGT 720

GATTTCACAG GCTCAGTGCT ACGGCTGGAT GTGGACACAG ACATGTGCAA CGTGCCTTAT 780

TCCATACCAA GGAGCAACCC ACACTTCAAC AGCACCAACC AGCCCCCCGA AGTGTTTGCT 840

CATGGGCTCC ACGATCCAGG CAGATGTGCT GTGGATAGAC ATCCCACTGA TATAAACATC 900

AATTTAACGA TACTGTGTTC AGACTCCAAT GGAAAAAACA GATCATCAGC CAGAATTC 958

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 597 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GAATTCAAGC CATTCAGTAA TGGTCCTTTG GTTGGTGGAT TTGTATACCG GGGCTGCCAG 60

TCAGAAAGAT TGTATGGAAG CTACGTGTTT GGAGATCGTA ATGGGAATTT CCTAACTCTC 120

CAGCAAAGTC CTGTGACAAA GCAGTGGCAA GAAAAACCAC TCTGTCTCGG CACTAGTGGG 180

TCCTGTAGAG GCTACTTTTC CGGTCACATC TTGGGATTTG GAGAAGATGA ACTAGGTGAA 240

GTTTACATTT TATCAAGCAG TAAAAGTATG ACCCAGACTC ACAATGGAAA ACTCTACAAA 300

ATTGTAGATC CCAAAAGACC TTTAATGCCT GAGGAATGCA GAGCCACGGT ACAACCTGCA 360

CAGACACTGA CTTCAGAGTG CTCCAGGCTC TGTCGAAACG GCTACTGCAC CCCCACGGGA 420

AAGTGCTGCT GCAGTCCAGG CTGGGAGGGG GACTTCTGCA GAACTGCAAA ATGTGAGCCA 480

GCATGTCGTC ATGGAGGTGT CTGTGTTAGA CCGAACAAGT GCCTCTGTAA AAAAGGATAT 540

CTTGGTCCTC AATGTGAACA AGTGGACAGA AACATCCGCA GAGTGACCAG GGCAGGT 597

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 426 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

ATGCTCAAGA TGCTGCCGTT CAAGCTGCTG CTGGTGGCCG TGGCTCTGTG CTTCTTCGAG 60

GGGGATGCCA AGTTCGGGGA GAGCGGCGCG CGGAGGAGAA GGTGCCTCAA CGGGACCCCC 120

GCGGCGGCTG AAGAAGCGCG ACCGGCGGCT GCTGTCCCCG GACCGGGCGG CGCGGAGGCG 180

ATGTGCCGCG GCCTCTACCC GCGCCTCTCC TGCTGCTCCC CGGCCGACGC GCAGGGGTTG 240

CTGCACGCCG GGGCCAAGAT ACTTTCTGTC ACGAACAACA CAGAATGTGC GAAGCTACTG 300

GAGGAAATCA AATGCGCACA CTGCTCACCT CATGCCCAGA ATCTTTTCCA CTCACCTGAG 360

AAAGGGGAAA CTTCTGAAAG AGAACTAACT CTTCCCTACT TGTGCAAAGA CTATTGTAAA 420

GAATTC 426

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1011 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GAATTCTATT ATACTTGCAG AGGTCACTTA CCAGGTTTTC TCCAAACTAC AGCTGATGAG 60

TTTTGCTTTT ACTATGCAAG AAAAGATGGT GGTGTATGCT TTCCAGATTT TCCAAGAAAA 120

CAAGTGCGAG GGCCAGCTTC TAACTCCCTG GACCACATGG AGGAATATGA CAAAGAGGAA 180

GAGATCAGCA GAAAGCACAA GCACAACTGC TTCTGTATTC AGGAAGTCAT GAGCGGACTA 240

AGGCAGCCTG TTGGAGCGGT ACATTGTGGG GATGGATCTC ATCGCCTCTT TATTCTTGAG 300

AAAGAAGGAT ATGTGAAGAT TTTCAGTCCT GAAGGAGACA TGATCAAGGA ACCTTTTTTG 360

GATATACACA AGCTTGTTCA AAGTGGAATA AAGGGAGGAG ATGAAAGAGG ACTGTTAAGC 420

CTTGCATTCC ATCCCAATTA CAAGAAAAAT GGAAAGCTGT ATGTGTCTTA TACCACCAAC 480

CAAGAACGGT GGGCTATTGG ACCTCATGAT CACATCCTTA GGGTGGTAGA ATACACAGTA 540

TCCAGGAAAA ATCCACAACA AGTTGATATA AGAACAGCCA GAGTGTTTTT AGAAGTAGCA 600

GAACTACATC GAAAACATCT AGGAGGGCAG CTTCTGTTTG GCCCAGATGG TTTCTTATAC 660

GTTTTCCTTG GAGATGGCAT GATTACCCTC GACGATATGG AAGAAATGGA TGGTTTAAGC 720

GATTTTACAG GTTCTGTATT ACGCCTCGAT GTAAATACTG ACCTGTGCAG TGTCCCTTAT 780

TCCATACCAC GGAGCAACCC ACATTTTAAT AGCACAAACC AACCTCCTGA AATTTTTGCA 840

CACGGACTCC ACAATCCAGG CCGATGTGCT GTGGATCACC ACCCAGCAGA TGTAAACATC 900

AATTTAACAA TACTTTGCTC AGATTCAAAT GGAAAGAACA GATCTTCAGC AAGAATCTTA 960

CAGATAATAA AGGGTAAAGA CTATGAAAGT GAGCCTTCAC TTTTAGAATT C 1011

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 660 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GAATTCAAAC CATTCAGCAG TGGAGCGTTG GTCGGTGGAT TTGTCTATCG AGGTTGCCAG 60

TCTGAAAGGC TCTACGGAAG TTATGTATTT GGAGACCGCA ATGGAAATTT TTTAACGCTG 120

CAACAGAATC CTGCAACTAA ACAGTGGCAA GAGAAACCCC TCTGTCTTGG CAACAGCGGT 180

TCATGTAGAG GTTTCTTTTC AGGCCCTGTC TTGGGATTTG GTGAAGATGA ACTAGGCGAG 240

ATTTACATAT TATCAAGCAG TAAAAGTATG ACACAGACTC ACAATGGAAA ACTCTACAAG 300

ATCATTGACC CAAAAAGGCC TTTAGTTCCT GAAGAATGCA AAAGAACAGC TCGGTCGGCA 360

CAGATACTGA CATCTGAATG CTCAAGGCAC TGCCGGAATG GGCACTGCAC ACCCACAGGA 420

AAATGCTGCT GTAATCAAGG CTGGGAAGGA GAGTTCTGCA GAACTGCAAA GTGTGACCCA 480

GCATGTCGAC ATGGAGGTGT CTGTGTAAGG CCTAATAAAT GCTTATGTAA AAAAGGCTAT 540

CTTGGCCCCC AGTGTGAACA ATTGGATTTA AACTTCCGAA AAGTTACAAG GCCAGGTATT 600

CTTGATCAGA TCCTAAACAT GACATCCTAC TTGCTGGATC TAACCAGCTA TATTGTATAG 660

We Claims

1. An isolated and/or recombinantly produced *HIP* polypeptide.

2. An isolated and/or recombinantly produced mammalian *HIP* polypeptide.

3. An isolated and/or recombinantly produced human *HIP* polypeptide.

4. An isolated and/or recombinantly produced polypeptide comprising a *HIP* amino acid sequence that binds to a *hedgehog* protein.

5. The polypeptide of claim 4, wherein the *HIP* amino acid sequence can be encoded by a nucleic acid which hybridizes under stringent conditions to a sequence selected from the group consisting of SEQ ID. Nos. 1, 2, 3, 4, 9, 10, 11, 12, 13 and 14.

6. The polypeptide of claim 4, which *HIP* amino acid sequence is at least 60% identical with a sequence selected from the group consisting of SEQ ID. Nos. 5, 6, 7 and 8, or a portion thereof.

7. The polypeptide of claims 4 or 6, wherein the *HIP* amino acid sequence is at least 25 amino acid residues in length.

8. An isolated and/or recombinantly produced polypeptide comprising a *HIP* amino acid sequence sufficient to bind to a *hedgehog* protein, which *HIP* amino acid sequence is at least 60% identical with a sequence selected from the group consisting of residues 18-678 of SEQ ID. No. 5 and residues 18-678 of SEQ ID. No. 6.

9. The polypeptide of any of claims 1-8, which polypeptide (i) regulates differentiation of neuronal cells, (ii) regulates survival of differentiated neuronal cells, (iii) regulates proliferation of chondrocytes, (iv) regulates proliferation of testicular germ line cells, and/or (v) regulates expression of a *patched* or *hedgehog* gene.

10. The polypeptide of any of claims 1-8, which polypeptide is a fusion protein.

11. The polypeptide of any of claims 1-8, wherein the polypeptide promotes differentiation of neuronal cells or survival of differentiated neuronal cells.

12. The polypeptide of claim 11, wherein the neuronal cell is a dopaminergic neuron.

13. The polypeptide of claim 11, wherein the neuronal cell is a motorneuron.

14. The polypeptide of any of claims 1-8, wherein the polypeptide regulates proliferation of chondrocytes.

15. The polypeptide of any of claims 1-8, wherein the polypeptide regulates spermatogenesis.

16. The polypeptide of claim 6, wherein the *HIP* amino acid sequence is at least 70 percent identical to a sequence represented in one of SEQ ID No:5, SEQ ID No:6, SEQ ID No:7 and SEQ ID No:8

17. The polypeptide of claim 6, wherein the *HIP* amino acid sequence is at least 80 percent identical to a sequence represented in one of SEQ ID No:5, SEQ ID No:6, SEQ ID No:7 and SEQ ID No:8

18. The polypeptide of claim 6, wherein the *HIP* amino acid sequence is at least 90 percent identical to a sequence represented in one of SEQ ID No:5, SEQ ID No:6, SEQ ID No:7 and SEQ ID No:8

19. The polypeptide of claim 6, wherein the *HIP* amino acid sequence is at least 95 percent identical to a sequence represented in one of SEQ ID No:5, SEQ ID No:6, SEQ ID No:7 and SEQ ID No:8

20. The polypeptide of claim 6, wherein the *HIP* amino acid sequence is identical to a sequence represented in one of SEQ ID No:5, SEQ ID No:6, SEQ ID No:7 and SEQ ID No:8

21. The polypeptide of claim 4, wherein the *HIP* amino acid sequence is encoded by a nucleic acid which hybridizes to the nucleic acid of SEQ ID No. 2.

22. The polypeptide of claim 4, wherein the *HIP* amino acid sequence is encoded by a naturally occurring *hedgehog* gene of a mammal.

23. The polypeptide of claim 4, wherein the *HIP* amino acid sequence is encoded by a naturally occurring *hedgehog* gene of a human.

24. The polypeptide of claim 7, wherein the *HIP* amino acid sequence corresponds to a fragment of at least 100 amino acid residues of a core polypeptide sequence of the *HIP* protein.

25. An isolated and/or recombinantly produced polypeptide comprising a *HIP* amino acid sequence immunologically crossreactive with an antibody which specifically binds a *HIP* protein having an amino acid sequence selected from the group consisting of SEQ ID No:1, SEQ ID No:2, SEQ ID No:3 and SEQ ID No:4, which *HIP* amino acid sequence binds to a *hedgehog* protein.

26. An isolated and/or recombinantly produced antibody or antibody fragment which is specifically immunoreactive with a *HIP* protein.

27. A monoclonal antibody specifically immunoreactive with a *HIP* protein.

28. A hybridoma producing the antibody of claim 27.

29. An isolated nucleic acid comprising coding sequence encoding a *HIP* polypeptide.

30. An isolated nucleic acid comprising *HIP* coding sequence encoding a *HIP* amino acid sequence that binds a *hedgehog* protein.

31. The nucleic acid of claim 30, wherein the *HIP* amino acid sequence is characterized by one or more of (i) the amino acid sequence is at least 60% identical with a sequence selected from the group consisting of SEQ ID. Nos. 5, 6, 7 and 8, and (ii) the *HIP* coding sequence hybridizes under stringent conditions to a sequence selected from the group consisting of SEQ ID. Nos. 1, 2, 3, 4, 9, 10, 11, 12, 13 and 14.

32. An nucleic acid comprising (i) a coding sequence of claim 30, and (ii) a heterologous transcriptional regulatory sequence.

33. The nucleic acid of claim 31, wherein the *HIP* coding sequence is from a naturally occurring *hedgehog* gene of a mammal.

34. The nucleic acid of claim 33, wherein the *HIP* gene is a human *HIP* gene.

35. The nucleic acid of claim 30, wherein the *HIP* amino acid sequence corresponds to an extracellular fragment of a *HIP* protein.

36. An expression vector, capable of replicating in at least one of a prokaryotic cell and eukaryotic cell, comprising the nucleic acid of claim 30 or 31.

37. A host cell transfected with the expression vector of claim 36 and expressing said recombinant polypeptide.

38. A method of producing a recombinant *HIP* polypeptide comprising culturing the cell of claim 37 in a cell culture medium to express said *HIP* polypeptide and isolating said *HIP* polypeptide from said cell culture.

39. A recombinant transfection system, comprising

(i) a gene construct including the nucleic acid of claim 30 or 31 operably linked to a transcriptional regulatory sequence for causing expression of the *HIP* polypeptide in eukaryotic cells, and

(ii) a gene delivery composition for delivering said gene construct to a cell and causing the cell to be transfected with said gene construct.

40. The recombinant transfection system of claim 39, wherein the gene delivery composition is selected from a group consisting of a recombinant viral particle, a liposome, and a poly-cationic nucleic acid binding agent.

41. A probe/primer comprising a substantially purified oligonucleotide, said oligonucleotide containing a region of nucleotide sequence which hybridizes under stringent conditions to at least 10 consecutive nucleotides of sense or antisense sequence of SEQ ID No. 1, 2, 3, 4, 9, 10, 11, 12, 13 or 14, or naturally occuring mutants thereof.

42. The probe/primer of claim 41, wherein the oligonucleotide further comprises a label group attached thereto and able to be detected.

43. A test kit for detecting cells which contain a *HIP* mRNA transcript, comprising a probe/primer of claim 41.

44. A purified preparation of an antisense nucleic acid which specifically hybridizes to and inhibits expression of a *HIP* gene under physiological conditions, which nucleic acid is at least one of (i) a synthetic oligonucleotide, (ii) single-stranded, (iii) linear,

(iv) 10 to 50 nucleotides in length. and (v) a DNA analog resistant to nuclease degradation.

45. The preparation of claim 44, which antisense nucleic acid is a DNA analog resistant to nuclease degradation.

46. A transgenic animal having cells which harbor a transgene comprising the nucleic acid of claim 29.

47. A transgenic animal in which *HIP*-dependent signal transduction pathways are inhibited in one or more tissue of the animal by one of either expression of an antagonistic *HIP* polypeptide or disruption of a *HIP* gene.

48. A method for modulating cell growth, differentiation or survival in an animal, comprising administering a therapeutically effective amount of an agent which induces, potentiates or inhibits *HIP*-dependent signal transduction.

49. The method of claim 48, comprising administering a nucleic acid construct encoding a *HIP* polypeptide under conditions wherein the construct is incorporated and recombinantly expressed by the cells to be modulated or cells located proximate thereto.

50. The method of claim 48, comprising administering an agent that inhibits interaction of *hedgehog* proteins with a *HIP* protein.

51. The method of claim 50, wherein the agent is a small organic molecule.

52. The method of claim 50, wherein the agent is a soluble extracellular domain of a *HIP* protein which binds to the *hedgehog* protein.

53. A method for determining if a subject is at risk for a disorder characterized by unwanted cell proliferation, differentiation or death, comprising detecting, in a tissue of the subject, the presence or absence of a genetic lesion characterized by at least one of (i) a mutation of a gene encoding a *HIP* protein; and (ii) the mis-expression of the gene.

54. The method of claim 53, wherein detecting the genetic lesion comprises ascertaining the existence of at least one of

i. a deletion of one or more nucleotides from the gene,
ii. an addition of one or more nucleotides to the gene,
iii. an substitution of one or more nucleotides of the gene,
iv. a gross chromosomal rearrangement of the gene,
v. aberrant methylation of the gene,
vi. a gross alteration in the level of a messenger RNA transcript of the gene,
vii. the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene, and
viii. a non-wild type level of the protein.

55. The method of claim 53, wherein detecting the genetic lesion comprises
i. providing a nucleic acid comprising an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence of SEQ ID No. 1, 2, 3, 4, 9, 10, 11, 12, 13 or 14, or naturally occurring mutants thereof or 5' or 3' flanking sequences naturally associated with the gene;
ii. exposing the nucleic acid to nucleic acid of the tissue; and
iii. detecting, by hybridization of the nucleic acid to the nucleic acid, the presence or absence of the genetic lesion.

56. The method of claim 54, wherein detection of the genetic lesion comprises detecting the presence or absence of a *HIP* protein on cells of a tissue sample and/or as soluble proteins in bodily fluid.

57. A method of detecting the presence of a *HIP* ligand on cells present in a biological sample, comprising contacting the cells with a labeled *HIP* polypeptide and under conditions where the *HIP* polypeptide can specifically bind to cognate ligand, and detecting presence of the *HIP* polypeptide bound to the cells.

58. An assay for identifying compounds which modulate *HIP* bioactivity, comprising:
(a) forming a reaction mixture including:
(i) a *HIP* polypeptide,
(ii) a *HIP* ligand, and
(iii) a test compound; and
(b) detecting interaction of the *HIP* polypeptide and ligand;
wherein a change in the interaction of the ligand and *HIP* polypeptide in the presence of the test compound, relative to the interaction in the absence of the test compound, indicates a potential *HIP* modulating activity for the test compound.

59. The assay of claim 58, wherein the reaction mixure is a cell-free protein preparation.

60. The assay claim 58, wherein the reaction mixure comprises a recombinant cell including a heterologous nucleic acid recombinantly expressing the *HIP* polypeptide.

61. The assay of claim 58, wherein the step of detecting interaction of the ligand and *HIP* polypeptide comprises a competitive binding assay.

62. The assay of claim 60, wherein the step of detecting interaction of the ligand and *HIP* polypeptide comprises detecting change in the level of an intracellular second messanger responsive to signalling by interaction of the ligand and *HIP* polypeptide.

63. The assay of claim 60, wherein the step of detecting interaction the ligand and *HIP* polypeptide comprises detecting change in the level of expression of a gene controlled by a transcriptional regulatory sequence responsive to *HIP*-dependent signal transduction.

64. The assay of claim 60 wherein the recombinant cell substantially lacks expression of an endogenous HIP protein.

65. The assay of claim 60 wherein the recombinant cell co-expresses a *patched* protein.

66. The assay of claim 60 wherein the recombinant cell co-expresses a *smoothened* protein.

67. The assay of claim 58, wherein the reaction mixture is a cell membrane preparation.

68. The assay of claim 58, wherein the reaction mixture is a reconstituted protein mixture.

69. The assay of claim 58, wherein the reaction mixture is a liposome reconsituting the *HIP* polypeptide as a *hedgehog* receptor.

70. The assay of claim 58, wherein the steps of the assay are repeated for a variegated library of at least 100 different test compounds.

71. The assay of claim 58, wherein the test compound is selected from the group consisting of small organic molecules, and natural product extracts.

72. The assay of claim 58, further comprising a step of preparing a pharmaceutical preparation of one or more compounds identified.

73. A compound which can be identified according to the assay of claim 58.

5

HIP-1 PROTEIN

```
human HIP-1          MLKMLSFKLL LLAVALGFFE GDAKFGERNE GSGARRRRCL NGNPPKRLKR   50
mouse HIP-1          MLKMLSFKLL LLAVALGFFE GDAKFGERSE GSGARRRRCL NGNPPKRLKR   50
chick HIP-1          MLKMLPFKLL LVAVALCFFE GDAKFGE--- -SGARRRRCL NGTPPRRLKK   46
zebrafish HIP-1      ---------- ---------- ---------- ---------- ----------
CONSENSUS            MLKML.FKLL L.AVAL.FFE GDAKFGE... .SGARRRRCL NG.PP.RLK.   50

human HIP_1          RDRRMMSQLE LLSGGEMLCG GFYPRLSCCL RSDSPGLGRL ENKIFSVTNN  100
mouse HIP-1          RDRRVMSQLE LLSGGEILCG GFYPRVSCCL QSDSPGLGRL ENKIFSATNN  100
chick HIP-1          RDRRLLSP-E APGGAEAMCR GLYPRLSCCS RADAQGLLHA GAKILSVTNN   95
zebrafish HIP-1      ---------- ---------- ---------- ---------- ----------
CONSENSUS            RDRR..S..E ...G.E..C. G.YPR.SCC. ..D..GL... ..KI.S.TNN  100

human HIP-1          TECGKLLEEI KCALCSPHSQ SLFHSPER-E VLERDIVLPL LCKDYCKEFF  149
mouse HIP-1          SECSRLLEEI QCAPCSPHSQ SLFYTPER-D VLDGDLALPL LCKDYCKEFF  149
chick HIP-1          TECAKLLEEI KCAHCSPHAQ NLFHSPEKGE TSERELTLPY LCKDYCKEFY  145
zebrafish HIP-1      ---------- ---------- ---------- ---------- ----------
CONSENSUS            .EC..LLEEI .CA.CSPH.Q .LF..PE.-. .......LP. LCKDYCKEF.  150

human HIP-1          YTCRGHIPGF LQTTADEFCF YYARKDGGLC FPDFPRKQVR GPASNYLDQM  199
mouse HIP-1          YTCRGHIPGL LQTTADEFCF YYARKDAGLC FPDFPRKQVR GPASNYLGQM  199
chick HIP-1          YTCRGHLPGF LQTTADEFCF YYARKDGGVC FPDFPRKQVR GPASNSLDHM  195
zebrafish HIP-1      ---------- ---------- ---------- ---------- ----------
CONSENSUS            YTCRGH.PG. LQTTADEFCF YYARKD.G.C FPDFPRKQVR GPASN.L..M  200

human HIP-1          EEYDKVEEIS RKHKHNCFCI QEVVSGLRQP VGALHSGDGS QRLFILEKEG  249
mouse HIP-1          EDYEKVGGIS RKHKHNCLCV QEVMSGLRQP VSAVHSGDGS HRLFILEKEG  249
chick HIP-1          EEYDKEEEIS RKHKHNCFCI QEVMSGLRQP VGAVHCGDGS HRLFILEKEG  245
zebrafish HIP-1      ---------- ---------- QEIHSGLQQP VGVVHCGDGS QRLFILEREG   30
CONSENSUS            E.Y.K...IS RKHKHNC.C. QEV.SGLRQP VGAVH.GDGS .RLFILEKEG  250
```

FIG. 1A

HIP-1 PROTEIN

```
human HIP-1     YVKILTPEGE IFKEPYLDIH KLVQSGIKGG DERGLLSLAF HPNYKKNGKL 299
mouse HIP-1     YVKILTPEGE LFKEPYLDIH KLVQSGIKGG DERGLLSLAF HPNYKKNGKL 299
chick HIP-1     YVKIFSPEGD MIKEPFLDIH KLVQSGIKGG DERGLLSLAF HPNYKKNGKL 295
zebrafish HIP-1 FVWILTHDME LLKEPFLDIH KLVQSGLKGG DERGLLSLAF HPNYKKNGKL  80
CONSENSUS       YVKILTPEGE ..KEP.LDIH KLVQSGIKGG DERGLLSLAF HPNYKKNGKL 300

human HIP_1     YVSYTTNQER WAIGPHDHIL RVVEYTVSRK NPHQVDLRTA RIFLEVAELH 349
mouse HIP-1     YVSYTTNQER WAIGPHDHIL RVVEYTVSRK NPHQVDVRTA RVFLEVAELH 349
chick HIP-1     YVSYTTNQER WAIGPHDHIL RVVEYTVSRK NPQQVDIRTA RVFLEVAELH 345
zebrafish HIP-1 YVSYTTNQER WTIGPHDHIL RVVEYTVSRK NPNQVDTRTP RVLMEVAELH 130
CONSENSUS       YVSYTTNQER WAIGPHDHIL RVVEYTVSRK NP.QVD.RTA RVFLEVAELH 350

human HIP-1     RKHLGGQLLF GPDGFLYIIL GDGMITLDDM EEMDGLSDFT GSVLRLDVDT 399
mouse HIP-1     RKHLGGQLLF GPDGFLYIIL GDGMITLDDM EEMDGLSDFT GSVLRLDVDT 399
chick HIP-1     RKHLGGQLLF GPDGFLYVFL GDGMITLDDM EEMDGLSDFT GSVLRLDVNT 395
zebrafish HIP-1 RKHLGGQLLF GPDGLLHIFL GDGMITLDNM EEMDGLSDFT GSVLRVDVDT 180
CONSENSUS       RKHLGGQLLF GPDGFLYI.L GDGMITLDDM EEMDGLSDFT GSVLRLDVDT 400

human HIP-1     DMCNVPYSIP RSNPHFNSTN QPPEVFAHGL HDPGRCAVDR HPTDININLT 449
mouse HIP-1     DMCNVPYSIP RSNPHFNSTN QPPEVFAHGL HDPGRCAVDR HPTDININLT 449
chick HIP-1     DLCSVPYSIP RSNPHFNSTN QPPEIFAHGL HNPGRCAVDH HPADVNINLT 445
zebrafish HIP-1 ECCSTPYSIP RNNPYFNSTN QPPEIFAHGL HDPGRCAVDK LRMDTNGSLL 230
CONSENSUS       D.C.VPYSIP RSNPHFNSTN QPPE.FAHGL HDPGRCAVD. HP.D.NINLT 450

human HIP-1     ILCSDSNGKN RSSARILQII KGKDYESEPS LLEFKPFSNG PLVGGFVYRG 499
mouse HIP-1     ILCSDSNGKN RSSARILQII KGRDYESEPS LLEFKPFSNG PLVGGFVYRG 499
chick HIP-1     ILCSDSNGKN RSSARILQII KGKDYESEPS LLEFKPFSSG ALVGGFVYRG 495
zebrafish HIP-1 ILCTDTVGKN TTTGRILQVI KGKDYENEPS MFDLGSSGGT TPVGGFIYRG 280
CONSENSUS       ILCSDSNGKN RSSARILQII KGKDYESEPS LLEFKPFS.G .LVGGFVYRG 500
```

FIG. 1B

HIP-1 PROTEIN

human HIP-1	CQSERLYGSY VFGDRNGNFL TLQQSPVTKQ WQEKPLCLGT SGSCRGYFSG	549
mouse HIP-1	CQSERLYGSY VFGDRNGNFL TLQQSPVTKQ WQEKPLCLGA SSSCRGYFSG	549
chick HIP-1	CQSERLYGSY VFGDRNGNFL TLQQNPATKQ WQEKPLCLGN SGSCRGFFSG	545
zebrafish HIP-1	CQSRRLYGSY VFGDKNGNFR ILQRPLEDRL WQEKPLCLGT SSSCGSSLVG	330
CONSENSUS	CQSERLYGSY VFGDRNGNFL TLQQ.P.TKQ WQEKPLCLG. S.SCRG.FSG	550
human HIP_1	HILGFGEDEL GEVYILSSSK SMTQTHNGKL YKIVDPKRPL MPEECRATVQ	599
mouse HIP-1	HILGFGEDEL GEVYILSSSK SMTQTHNGKL YKIVDPKRPL MPEECRVTVQ	599
chick HIP-1	PVLGFGEDEL GEIYILSSSK SMTQTHNGKL YKIIDPKRPL VPEECKRTAR	595
zebrafish HIP-1	HILGFGEDEL GEVYILVSSK STAKQSHGKI YKLVDPKRPQ VPKECRRPVE	380
CONSENSUS	HILGFGEDEL GEVYILSSSK SMTQTHNGKL YKIVDPKRPL .PEECR.TV.	600
human HIP-1	PAQTLTSECS RLCRNGYCTP TGKCCCSPGW EGDFCRTAKC EPACRHGGVC	649
mouse HIP-1	PAQPLTSDCS RLCRNGYYTP TGKCCCSPGW EGDFCRIAKC EPACRHGGVC	649
chick HIP-1	SAQILTSECS RHCRNGHCTP TGKCCCNQGW EGEFCRTAKC DPACRHGGVC	645
zebrafish HIP-1	DPEMLSTACS RECKNGHCTP TGKCCCNAGW EGPFCLRAKC ELACRNGGVC	430
CONSENSUS	.AQ.LTS.CS R.CRNG.CTP TGKCCC..GW EG.FCR.AKC EPACRHGGVC	650
human HIP-1	VRPNKCLCKK GYLGPQCEQV DRNIR-RMTR AGVLDQIFDM TSYLLDLTNY	698
mouse HIP-1	VRPNKCLCKK GYLGPQCEQV DRNVR-RVTR AGILDQIIDM TSYLLDLTSY	698
chick HIP-1	VRPNKCLCKK GYLGPQCEQV DRNFR-KVTR PGILDQILDM TSYLLDLTSY	694
zebrafish HIP-1	VEPNKCLCKE GFSGNQCSKG ERGTKGDGEK DSILEHIIDM TTYLLDLTSY	480
CONSENSUS	VRPNKCLCKK GYLGPQCEQV DRN.R-..TR .GILDQI.DM TSYLLDLTSY	700
human HIP-1	IV	700
mouse HIP-1	IV	700
chick HIP-1	IV	696
zebrafish HIP-1	IV	482
CONSENSUS	IV	702

FIG. 1C

```
human HIP-1       ATGCTGAAGA TGCTCTCCTT TAAGCTGCTG CTGCTGGCCG TGGCTCTGGG  50
mouse HIP-1       ATGCTGAAGA TGCTCTCGTT TAAGCTGCTA CTGCTGGCCG TGGCTCTGGG  50
chick HIP-1       ATGCTCAAGA TGCTGCCGTT CAAGCTGCTG CTGGTGGCCG TGGCTCTGTG  50
zebrafish HIP-1   ---------- ---------- ---------- ---------- ----------
CONSENSUS         ATGCTSAAGA TGCTSYCSTT YAAGCTGCTR CTGSTGGCCG TGGCTCTGKG  50

human HIP_1       CTTCTTTGAA GGAGATGCTA AGTTTGGGGA AAGAAACGAA GGGAGCGGAG 100
mouse HIP-1       CTTCTTTGAA GGAGATGCGA AGTTTGGGGA AAGGAGCGAG GGGAGCGGAG 100
chick HIP-1       CTTCTTCGAG GGGGATGCCA AGTTCGGGGA ---------- --GAGCGGCG  88
zebrafish HIP-1   ---------- ---------- ---------- ---------- ----------
CONSENSUS         CTTCTTYGAR GGRGATGCBA AGTTYGGGGA AAGRARCGAR GGGAGCGGMG 100

human HIP-1       CAAGGAGGAG AAGGTGCCTG AATGGGAACC CCCCGAAGCG CCTGAAAAGG 150
mouse HIP-1       CGAGAAGGAG ACGGTGCCTG AATGGGAACC CCCCAAAGCG CCTAAAGAGA 150
chick HIP-1       CGCGGAGGAG AAGGTGCCTC AACGGGACCC CGCCGCGGCG GCTGAAGAAG 138
zebrafish HIP-1   ---------- ---------- ---------- ---------- ----------
CONSENSUS         CRMGRAGGAG AMGGTGCCTS AAYGGGAMCC CSCCRMRGCG SCTRAARARR 150

human HIP-1       AGAGACAGGA GGATGATGTC CCAGCTGGAG CTGCTGAGTG GGGGAGAGAT 200
mouse HIP-1       AGGGACAGGC GGGTGATGTC CCAGCTGGAG CTGCTCAGTG GAGGAGAGAT 200
chick HIP-1       CGCGACCGGC GGCTGCTGTC C---CCGGAG GCGCCGGGCG GCGCGGAGGC 185
zebrafish HIP-1   ---------- ---------- ---------- ---------- ----------
CONSENSUS         MGVGACMGGM GGVTGMTGTC CCAGCYGGAG SYGCYSRGYG GVGSRGAGRY 200

human HIP-1       GCTGTGCGGT GGCTTCTACC CTCGGCTGTC CTGCTGCCTG CGGAGTGACA 250
mouse HIP-1       CCTGTGTGGT GGCTTCTACC CACGAGTATC TTGCTGCCTG CAGAGTGACA 250
chick HIP-1       GATGTGCCGC GGCCTCTACC CGCGCCTCTC CTGCTGCTCC CGCGCCGACG 235
zebrafish HIP-1   ---------- ---------- ---------- ---------- ----------
CONSENSUS         SMTGTGYSGY GGCYTCTACC CDCGVSTVTC YTGCTGCYYS CRSRSYGACR 250
```

HIP-1 cDNA

FIG. 1D

HIP-1 cDNA

```
human HIP-1          GCCCGGGGCT AGGGCGCCTG GAGAATAAGA TATTTTCTGT TACCAACAAC   300
mouse HIP-1          GCCCTGGATT GGGGCGTCTG GAGAACAAGA TCTTTTCTGC CACCAACAAC   300
chick HIP-1          CGCAGGGGTT GCTGCACGCC GGGGCCAAGA TACTTTCTGT CACGAACAAC   285
zebrafish HIP-1      ---------- ---------- ---------- ---------- ----------
CONSENSUS            SSCMKGGRYT RSKGCRYSYS GRGRMYAAGA TMYTTTCTGY YACSAACAAC   300

human HIP_1          ACAGAATGTG GGAAGTTACT GGAGGAAATC AAATGTGCAC TTTGCTCTCC   350
mouse HIP-1          TCAGAATGCA GCAGGCTGCT GGAGGAGATC CAATGTGCTC CCTGCTCCCC   350
chick HIP-1          ACAGAATGTG CGAAGCTACT GGAGGAAATC AAATGCGCAC ACTGCTCACC   335
zebrafish HIP-1      ---------- ---------- ---------- ---------- ----------
CONSENSUS            WCAGAATGYR SSARGYTRCT GGAGGARATC MAATGYGCWC HYTGCTCHCC   350

human HIP-1          ACATTCTCAA AGCCTGTTCC ACTCACCTGA GA---GAGAA GTCTTGGAAA   397
mouse HIP-1          GCATTCCCAG AGCCTCTTCT ACACACCTGA AA---GAGAT GTCCTGGATG   397
chick HIP-1          TCATGCCCAG AATCTTTTCC ACTCACCTGA GAAAGGGGAA ACTTCTGAAA   385
zebrafish HIP-1      ---------- ---------- ---------- ---------- ----------
CONSENSUS            DCATKCYCAR ARYCTBTTCY ACWCACCTGA RAAAGGRGAW RYYYYKGAWR   400

human HIP-1          GAGACATAGT ACTTCCTCTG CTCTGCAAAG ACTATTGCAA AGAATTCTTT   447
mouse HIP-1          GGGACCTAGC ACTTCCGCTC CTCTGCAAAG ACTACTGCAA AGAATTCTTT   447
chick HIP-1          GAGAACTAAC TCTTCCCTAC TTGTGCAAAG ACTATTGTAA AGAATTCTAT   435
zebrafish HIP-1      ---------- ---------- ---------- ---------- ----------
CONSENSUS            GRGAMMTARY WCTTCCBYWS YTSTGCAAAG ACTAYTGYAA AGAATTCTWT   450

human HIP-1          TACACTTGCC GAGGCCATAT TCCAGGTTTC CTTCAAACAA CTGCGGATGA   497
mouse HIP-1          TATACTTGCC GAGGCCATAT TCCAGGTCTT CTTCAAACAA CTGCTGATGA   497
chick HIP-1          TATACTTGCA GAGGTCACTT ACCAGGTTTT CTCCAAACTA CAGCTGATGA   485
zebrafish HIP-1      ---------- ---------- ---------- ---------- ----------
CONSENSUS            TAYACTTGCM GAGGYCAYWT WCCAGGTYTY CTYCAAACWA CWGCKGATGA   500
```

FIG. 1E

HIP-1 cDNA

```
human HIP-1       GTTTTGCTTT TACTATGCAA GAAAAGATGG TGGGTTGTGC TTTCCAGATT   547
mouse HIP-1       ATTTTGCTTT TACTATGCAA GAAAAGATGC TGGGTTATGC TTTCCAGACT   547
chick HIP-1       GTTTTGCTTT TACTATGCAA GAAAAGATGG TGGTGTATGC TTTCCAGATT   535
zebrafish HIP-1   ---------- ---------- ---------- ---------- ----------
CONSENSUS         RTTTTGCTTT TACTATGCAA GAAAAGATGS TGGKKTRTGC TTTCCAGAYT   550

human HIP_1       TTCCAAGAAA ACAAGTCAGA GGACCAGCAT CTAACTACTT GGACCAGATG   597
mouse HIP-1       TCCCGAGAAA GCAAGTCAGA GGACCAGCAT CTAACTACTT GGGCCAGATG   597
chick HIP-1       TTCCAAGAAA ACAAGTGCGA GGGCCAGCTT CTAACTCCCT GGACCACATG   585
zebrafish HIP-1   ---------- ---------- ---------- ---------- ----------
CONSENSUS         TYCCRAGAAA RCAAGTSMGA GGRCCAGCWT CTAACTMCYT GGRCCASATG   600

human HIP-1       GAAGAATATG ACAAAGTGGA AGAGATCAGC AGAAAGCACA AACACAACTG   647
mouse HIP-1       GAAGACTACG AGAAAGTGGG GGGGATCAGC AGAAAACACA AACACAACTG   647
chick HIP-1       GAGGAATATG ACAAAGAGGA AGAGATCAGC AGAAAGCACA AGCACAACTG   635
zebrafish HIP-1   ---------- ---------- ---------- ---------- ----------
CONSENSUS         GARGAMTAYG ASAAAGWGGR RGRGATCAGC AGAAARCACA ARCACAACTG   650

human HIP-1       CTTCTGTATT CAGGAGGTTG TGAGTGGGCT GCGGCAGCCC GTTGGTGCCC   697
mouse HIP-1       CCTCTGTGTC CAGGAGGTCA TGAGTGGGCT GCGGCAGCCT GTGAGCGCTG   697
chick HIP-1       CTTCTGTATT CAGGAAGTCA TGAGCGGACT AAGGCAGCCT GTTGGAGCGG   685
zebrafish HIP-1   ---------- CAGGAGATCC ATAGTGGTCT TCAACAACCT GTTGGCGTGG    40
CONSENSUS         CYTCTGTRTY CAGGARRTYV WKAGYGGDCT DMRRCARCCY GTKRGHGYBS   700

human HIP-1       TGCATAGTGG GGATGGCTCG CAACGTCTCT TCATTCTGGA AAAAGAAGGT   747
mouse HIP-1       TGCACAGCGG GGATGGCTCC CATCGGCTCT TCATTCTAGA GAAGGAAGGC   747
chick HIP-1       TACATTGTGG GGATGGATCT CATCGCCTCT TTATTCTTGA GAAAGAAGGA   735
zebrafish HIP-1   TGCATTGTGG AGATGGATCG CAGCGGCTTT TTATATTGGA GAGGGAAGGC    90
CONSENSUS         TRCAYWGYGG RGATGGMTCB CADCGBCTYT TYATWYTDGA RARRGAAGGH   750
```

FIG. 1F

```
human HIP-1       TATGTGAAGA TACTTACCCC TGAAGGAGAA ATTTTCAAGG AGCCTTATTT  797
mouse HIP-1       TACGTGAAAA TTCTAACCCC AGAAGGAGAA CTGTTCAAGG AGCCTTACTT  797
chick HIP-1       TATGTGAAGA TTTTCAGTCC TGAAGGAGAC ATGATCAAGG AACCTTTTTT  785
zebrafish HIP-1   TTTGTGTGGA TCCTCACACA TGACATGGAA CTCCTAAAAG AGCCTTTTCT  140
CONSENSUS         TWYGTGWRRA THYTHASHCM WGAMRKRGAM MTBHTMAARG ARCCTTWYYT  800

human HIP-1       GGACATTCAC AAACTTGTTC AAAGTGGAAT AAAGGGAGGA GATGAAAGAG  847
mouse HIP-1       GGACATTCAC AAACTTGTTC AAAGTGGAAT AAAGGGAGGA GACGAAAGGG  847
chick HIP-1       GGATATACAC AAGCTTGTTC AAAGTGGAAT AAAGGGAGGA GATGAAAGAG  835
zebrafish HIP-1   GGACATTCAT AAGCTGGTAC AAAGTGGTTT AAAGGGGGGA GATGAAAGGG  190
CONSENSUS         GGAYATWCAY AARCTKGTWC AAAGTGGWWT AAAGGGRGGA GAYGAAAGRG  850

human HIP_1       GACTGCTAAG CCTCGCATTC CATCCCAATT ACAAGAAAAA TGGAAAGTTG  897
mouse HIP-1       GCCTGCTAAG CCTGGCATTC CATCCCAATT ACAAGAAAAA TGGAAAGCTG  897
chick HIP-1       GACTGTTAAG CCTTGCATTC CATCCCAATT ACAAGAAAAA TGGAAAGCTG  885
zebrafish HIP-1   GCTTGCTAAG CCTTGCATTC CACCCCAATT ATAAGAAAAA TGGCAAGCTC  240
CONSENSUS         GMYTGYTAAG CCTBGCATTC CAYCCCAATT AYAAGAAAAA TGGMAAGYTS  900

human HIP-1       TATGTGTCCT ATACCACCAA CCAAGAACGG TGGGCTATCG GGCCTCATGA  947
mouse HIP-1       TATGTGTCTT ATACCACCAA CCAGGAACGG TGGGCTATTG GGCCTCACGA  947
chick HIP-1       TATGTGTCTT ATACCACCAA CCAAGAACGG TGGGCTATTG GACCTCATGA  935
zebrafish HIP-1   TACGTCTCCT ATACGACCAA CCAGGAGCGA TGGACTATTG GACCACACGA  290
CONSENSUS         TAYGTSTCYT ATACSACCAA CCARGARCGR TGGRCTATYG GRCCWCAYGA  950

human HIP-1       CCACATTCTT AGGGTTGTGG AATACACAGT ATCCAGAAAA AATCCACACC  997
mouse HIP-1       CCACATTCTT CGGGTTGTGG AATACACAGT ATCCAGGAAA AACCCCCATC  997
chick HIP-1       TCACATCCTT AGGGTGGTAG AATACACAGT ATCCAGGAAA AATCCACAAC  985
zebrafish HIP-1   CCACATTCTT CGTGTAGTGG AGTACACAGT GTCCAGAAAA AATCCAAACC  340
CONSENSUS         YCACATYCTT MGKGTDGTRG ARTACACAGT RTCCAGRAAA AAYCCMMAHC 1000
```

HIP-1 cDNA

FIG. 1G

```
human HIP-1     AAGTTGATTT GAGAACAGCC AGAATCTTTC TTGAAGTTGC AGAACTCCAC 1047
mouse HIP-1     AAGTTGATGT GAGAACAGCC AGGGTGTTTC TGGAAGTCGC AGAGCTCCAC 1047
chick HIP-1     AAGTTGATAT AAGAACAGCC AGAGTGTTTT TAGAAGTAGC AGAACTACAT 1035
zebrafish HIP-1 AGGTGGACAC AAGGACTCCT CGGGTTTTAA TGGAAGTTGC AGAACTTCAC  390
CONSENSUS       ARGTKGAYDY RAGRACWSCY MGRRTBTTWH TDGAAGTHGC AGARCTHCAY 1050

human HIP-1     AGAAAGCATC TGGGAGGACA ACTGCTCTTT GGCCCTGACG GCTTTTTGTA 1097
mouse HIP-1     CGAAAGCATC TTGGGGGACA GCTGCTCTTT GGTCCTGATG GCTTTTTGTA 1097
chick HIP-1     CGAAAACATC TAGGAGGGCA GCTTCTGTTT GGCCCAGATG GTTTCTTATA 1085
zebrafish HIP-1 CGAAAGCATC TGGGAGGCCA GCTCCTCTTT GGGCCTGATG GGCTTCTGCA  440
CONSENSUS       MGAAARCATC TDGGRGGVCA RCTBCTSTTT GGBCCWGAYG GBYTYYTRYA 1100

human HIP_1     CATCATTCTT GGTGATGGGA TGATTACACT GGATGATATG GAAGAAATGG 1147
mouse HIP-1     CATCATCCTT GGGGATGGTA TGATCACATT GGATGACATG GAAGAGATGG 1147
chick HIP-1     CGTTTTCCTT GGAGATGGCA TGATTACCCT CGACGATATG GAAGAAATGG 1135
zebrafish HIP-1 CATCTTTTTA GGAGATGGCA TGATCACTTT GGACAATATG GAGGAGATGG  490
CONSENSUS       CRTYWTYYTW GGDGATGGBA TGATYACHYT SGAYRAYATG GARGARATGG 1150

human HIP-1     ATGGGTTAAG TGATTTCACA GGCTCAGTGC TACGGCTGGA TGTGGACACA 1197
mouse HIP-1     ATGGGTTAAG TGACTTCACA GGCTCTGTGC TGAGGCTGGA CGTGGACACC 1197
chick HIP-1     ATGGTTTAAG CGATTTTACA GGTTCTGTAT TACGCCTCGA TGTAAATACT 1185
zebrafish HIP-1 ATGGTCTGAG TGATTTCACA GGTTCTGTTC TTCGGGTGGA TGTGGACACA  540
CONSENSUS       ATGGKYTRAG YGAYTTYACA GGYTCWGTDY TDMGSSTSGA YGTRRAYACH 1200

human HIP-1     GACATG-TGC AACGTGCCTT ATTCCATACC AAGGAGCAAC CCACACTTCA 1246
mouse HIP-1     GACATG-TGC AATGTGCCTT ATTCCATACC TCGGAGTAAC CCTCACTTCA 1246
chick HIP-1     GACCTG-TGC AGTGTCCCTT ATTCCATACC ACGGAGCAAC CCACATTTTA 1234
zebrafish HIP-1 GA-ATGTTGT AGTACTCCCT ACTCCATACC CAGAAACAAT CCCTATTTCA  589
CONSENSUS       GACMTGTTGY ARYRYBCCYT AYTCCATACC HMGRARYAAY CCHYAYTTYA 1250
```

FIG. 1H

HIP-1 cDNA

HIP-1 cDNA

human HIP-1	CACAATGGAA AACTCTACAA AATTGTAGAT CCCAAAAGAC CTTTAATGCC	1772
mouse HIP-1	CACAATGGAA AACTCTACAA GATCGTAGAC CCCAAAAGAC CTTTAATGCC	1772
chick HIP-1	CACAATGGAA AACTCTACAA GATCATTGAC CCAAAAAGGC CTTTAGTTCC	1760
zebrafish HIP-1	CGC-ATGGAA AGATCTACAA GTTGGTGGAC CCCAAAAGAC CACAAGTTCC	1115
CONSENSUS	CRCAATGGAA ARMTCTACAA RWTBRTDGAY CCMAAAAGRC CWYWARTKCC	1800
human HIP_1	TGAGGAATGC AGAGCCACGG TACAACCTGC ACAGACACTG ACTTCAGAGT	1822
mouse HIP-1	TGAGGAATGC AGAGTCACAG TTCAACCTGC CCAGCCACTG ACCTCCGATT	1822
chick HIP-1	TGAAGAATGC AAAAGAACAG CTCGGTCGGC ACAGATACTG ACATCTGAAT	1810
zebrafish HIP-1	TAAGGAGTGC AGAAGACCAG TAGAAGATCC AGAGATGCTA AGCACTGCTT	1165
CONSENSUS	TRARGARTGC ARARBMMCRG YWSRRBMKSC MSAGMYRCTR ASHWCHGMDT	1850
human HIP-1	GCTCCAGGCT CTGTCGAAAC GGCTACTGCA CCCCCACGGG AAAGTGCTGC	1872
mouse HIP-1	GCTCCCGGCT CTGTCGAAAC GGCTACTACA CCCCCACTGG CAAGTGCTGC	1872
chick HIP-1	GCTCAAGGCA CTGCCGGAAT GGGCACTGCA CACCCACAGG AAAATGCTGC	1860
zebrafish HIP-1	GTTCACGTGA ATGCAAGAAC GGCCACTGTA CACCAACTGG CAAGTGCTGC	1215
CONSENSUS	GYTCMMGKSW MTGYMRRAAY GGSYACTRYA CMCCMACDGG MAARTGCTGC	1900
human HIP-1	TGCAGTCCAG GCTGGGAGGG GGACTTCTGC AGAACTGCAA AATGTGAGCC	1922
mouse HIP-1	TGCAGTCCCG GCTGGGAGGG AGACTTCTGC AGAATTGCCA AGTGTGAGCC	1922
chick HIP-1	TGTAATCAAG GCTGGGAAGG AGAGTTCTGC AGAACTGCAA AGTGTGACCC	1910
zebrafish HIP-1	TGCAATGCAG GCTGGGAAGG CCCCTTCTGC TTACGAGCCA AGTGTGAACT	1265
CONSENSUS	TGYARTSMMG GCTGGGARGG VSMSTTCTGC WKAMBWGCMA ARTGTGAVCY	1950
human HIP-1	AGCATGTCGT CATGGAGGTG TCTGTGTTAG ACCGAACAAG TGCCTCTGTA	1972
mouse HIP-1	AGCGTGCCGT CATGGAGGTG TCTGTGTCAG ACCGAACAAG TGCCTCTGTA	1972
chick HIP-1	AGCATGTCGA CATGGAGGTG TCTGTGTAAG GCCTAATAAA TGCTTATGTA	1960
zebrafish HIP-1	GGCTTGTCGC AATGGCGGGG TCTGTGTTGA GCCCAACAAG TGTCTCTGCA	1315
CONSENSUS	RGCDTGYCGH MATGGMGGKG TCTGTGTHRR RCCBAAYAAR TGYYTMTGYA	2000

FIG. 1K

HIP-1 cDNA

```
human HIP-1          AAAAAGGATA TCTTGGTCCT CAATGT-GAA CAAGTG-GAC AGAAACATCC 2020
mouse HIP-1          AAAAGGGCTA TCTTGGTCCT CAATGT-GAA CAAGTG-GAC AGGAACGTCC 2020
chick HIP-1          AAAAAGGCTA TCTTGGCCCC CAGTGT-GAA CAAGTG-GAT AGAAACTTCC 2008
zebrafish HIP-1      AGGAAGGTTT TTCTGGCAAC CAGTGCAGTA AAGGAGAGCG AGGGACAAAA 1365
CONSENSUS            ARRARGGHTW TYYTGGYMMY CARTGYAGWA MARGWGAGMB AGRRACDWMM 2050

human HIP_1          GCAGA--ATG ACCAGGGCAG GTGTTCTTGA TCAGATCTTC GACATGACAT 2068
mouse HIP-1          GCAGA--GTG ACCAGGGCAG GTATCCTTGA TCAGATCATT GACATGACGT 2068
chick HIP-1          GAAAA--GTT ACAAGGCCAG GTATTCTTGA TCAGATCCTA GACATGACAT 2056
zebrafish HIP-1      GGGGACGGTG AGAAAGACA- GCATCCTGGA GCACATCATT GACATGACGA 1414
CONSENSUS            GVRRACGRTK ASMARGVCAG GYRTYCTKGA KCASATCWTH GACATGACRW 2100

human HIP-1          CTTACTTGCT GGATCTAACA AATTACATTG TATAG                  2103
mouse HIP-1          CTTACTTGCT GGATCTCACA AGTTACATTG TATAG                  2103
chick HIP-1          CCTACTTGCT GGATCTAACC AGCTATATTG TATAG                  2091
zebrafish HIP-1      CTTACCTGCT GGACCTCACT AGTTATATTG TTTAA                  1449
CONSENSUS            CYTACYTGCT GGAYCTMACH ARYTAYATTG TWTAR                  2135
```

FIG. 1L

HIP-1 cDNA

human HIP-1	TTGGTGGATT TGTATACCGG GGCTGCCAGT CAGAAAGATT GTATGGAAGC	1524
mouse HIP-1	TTGGTGGATT TGTTTACAGA GGCTGTCAGT CTGAAAGATT GTACGGAAGC	1524
chick HIP-1	TCGGTGGATT TGTCTATCGA GGTTGCCAGT CTGAAAGGCT CTACGGAAGT	1512
zebrafish HIP-1	TTGGTGGATT TATCTACAGA GGATGTCAGT CAAGAAGACT TTACGGAAGT	867
CONSENSUS	TYGGTGGATT TRTHTAYMGR GGHTGYCAGT CWRRAAGRYT BTAYGGAAGY	1550
human HIP_1	TACGTGTTTG GAGATCGTAA TGGGAATTTC CTAACTCTCC AGCAAAGTCC	1574
mouse HIP-1	TATGTGTTCG GAGATCGCAA TGGGAATTTC TTAACCCTCC AGCAAAGCCC	1574
chick HIP-1	TATGTATTTG GAGACCGCAA TGGAAATTTT TTAACGCTGC AACAGAATCC	1562
zebrafish HIP-1	TATGTATTTG GAGACAAAAA TGGGAACTTT AGAATTCTCC AGAGGCCTTT	917
CONSENSUS	TAYGTRTTYG GAGAYMRHAA TGGRAAYTTY HKAAYBCTSC ARMRRMVYYY	1600
human HIP-1	TG-TGACAAA GCAGTGGCAA GAAAAACCAC TCTGTCTCGG CACTAGTGGG	1623
mouse HIP-1	AG-TGACCAA GCAATGGCAA GAAAAGCCGC TCTGCCTGGG TGCCAGCAGC	1623
chick HIP-1	TG-CAACTAA ACAGTGGCAA GAGAAACCCC TCTGTCTTGG CAACAGCGGT	1611
zebrafish HIP-1	AGAAGACCGA -TTGTGGCAA GAGAAGCCTC TTTGTCTTGG TACTAGCAGT	966
CONSENSUS	WGAHRACHRA RYWRTGGCAA GARAARCCNC TYTGYCTBGG YRMYAGYRGB	1650
human HIP-1	TCCTGTAGAG GCTACTTTTC CGGTCACATC TTGGGATTTG GAGAAGATGA	1673
mouse HIP-1	TCCTGTCGAG GCTACTTTTC GGGTCACATC TTGGGATTTG GAGAAGATGA	1673
chick HIP-1	TCATGTAGAG GTTTCTTTTC AGGCCCTGTC TTGGGATTTG GTGAAGATGA	1661
zebrafish HIP-1	TCCTGTGGTT CCTCGCTGGT AGGCCACATC CTGGGGTTTG GCGAAGATGA	1016
CONSENSUS	TCMTGTVGWK SYTHSYTKKY VGGYCMYRTC YTGGGRTTTG GHGAAGATGA	1700
human HIP-1	ACTAGGTGAA GTTTACATTT TATCAAGCAG TAAAAGTAT- GACCCAGACT	1722
mouse HIP-1	ATTAGGAGAG GTTTACATTC TATCAAGCAG TAAGAGTAT- GACCCAGACT	1722
chick HIP-1	ACTAGGCGAG ATTTACATAT TATCAAGCAG TAAAAGTAT- GACACAGACT	1710
zebrafish HIP-1	ATTAGGTGAG GTCTACATCC TTGTCTCCAG CAAGAGCACA GCCAAACAGT	1066
CONSENSUS	AYTAGGHGAR RTYTACATHY TWKYMWSCAG YAARAGYAYA GMCMMASAST	1750

FIG. 1J

HIP-1 cDNA

```
human HIP-1          ACAGCACCAA CCAGCCCCCC GAAGTGTTTG CTCATGGGCT CCACGATCCA 1296
mouse HIP-1          ACAGCACCAA CCAGCCCCCA GAAGTATTTG CCCACGGCCT CCATGATCCA 1296
chick HIP-1          ATAGCACAAA CCAACCTCCT GAAATTTTTG CACACGGACT CCACAATCCA 1284
zebrafish HIP-1      ACAGCACAAA TCAACCCCCC GAAATCTTTG CCCATGGTCT GCATGACCCA  639
CONSENSUS            AYAGCACMAA YCARCCYCCH GAARTNTTTG CHCAYGGNCT SCAYRAYCCA 1300

human HIP_1          GGCAGATGTG CTGTGGATAG ACATCCCACT GATATAAACA TCAATTTAAC 1346
mouse HIP-1          GGCAGATGTG CCGTGGATCG ACATCCTACT GATATAAACA TCAATTTAAC 1346
chick HIP-1          GGCCGATGTG CTGTGGATCA CCACCCAGCA GATGTAAACA TCAATTTAAC 1334
zebrafish HIP-1      GGGAGGTGTG CAGTAGATAA GCTCCGCATG GACACCAATG GGAGTCTGCT  689
CONSENSUS            GGSMGRTGTG CHGTRGATMR VCWYCSHRYD GAYRYMAAYR KSARTYTRMY 1350

human HIP-1          GATACTGTGT TCAGACTCCA ATGG-AAAAA ---------- ---------- 1375
mouse HIP-1          AATACTTTGC TCAGATTCCA ACGG-GAAAA ---------- ---------- 1375
chick HIP-1          AATACTTTGC TCAGATTCAA ATGG-AAAGA ---------- ---------- 1363
zebrafish HIP-1      GATCCTGTGC ACAGATACAG TTGGCAAAAA TACGACAACA GGCAGGATCC  739
CONSENSUS            RATMCTKTGY WCAGAYWCMR WYGGCRAARA TACGACAACA GGCAGGATCC 1400

human HIP-1          -ACAGATCAT CAGCCAGAAT TCTACAGATA ATAAAGGGGA AAGATTATGA 1424
mouse HIP-1          -ACAGGTCAT CAGCCAGAAT CCTACAGATA ATAAAGGGAA GAGATTATGA 1424
chick HIP-1          -ACAGATCTT CAGCAAGAAT CTTACAGATA ATAAAGGGTA AAGACTATGA 1412
zebrafish HIP-1      TACAGGTCAT CA-------- ---------- ---AA-GGGA AAGATTACGA  767
CONSENSUS            TACAGRTCWT CAGCMAGAAT YYTACAGATA ATAAAGGGDA RAGAYTAYGA 1450

human HIP-1          AAGTGAGCCA TCACTTTTAG AATTCAAGCC ATTCAGTAAT GGTCCTTTGG 1474
mouse HIP-1          AAGTGAGCCA TCTCTTCTTG AATTCAAGCC ATTCAGTAAC GGCCCTTTGG 1474
chick HIP-1          AAGTGAGCCT TCACTTTTAG AATTCAAACC ATTCAGCAGT GGAGCGTTGG 1462
zebrafish HIP-1      AAACGAGCCA TCTATGTTTG ACTTGGGGTC AAGCGGAGGT ACCACCCCTG  817
CONSENSUS            AARYGAGCCW TCWMTKYTWG AMTTSRRRYC AWKCRGHRRY RSHVCBYYKG 1500
```

FIG. 11

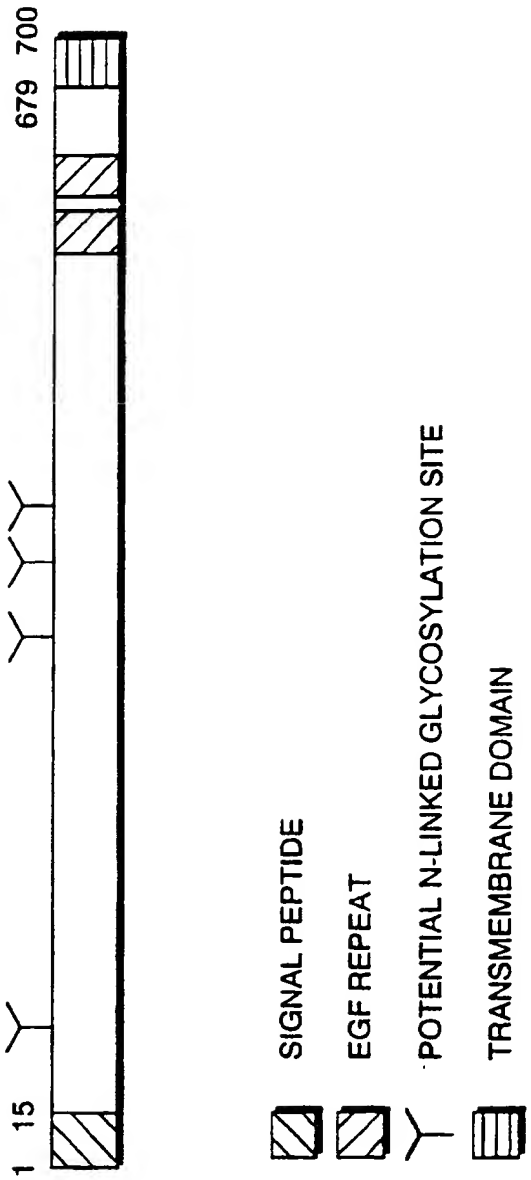

1
15
679
700
SIGNAL PEPTIDE
EGF REPEAT
POTENTIAL N-LINKED GLYCOSYLATION SITE
TRANSMEMBRANE DOMAIN

FIG. 2

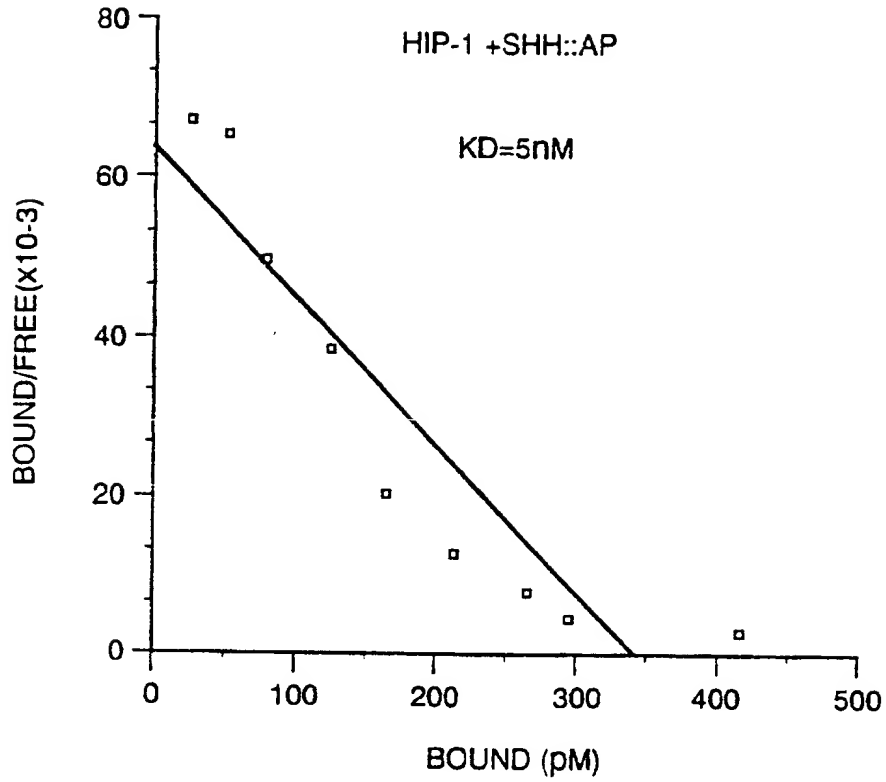

HIP-1 +SHH::AP
KD=5nM
BOUND/FREE(X10-3)
BOUND (pM)
0
20
40
60
80
0
100
200
300
400
500

FIG. 3A

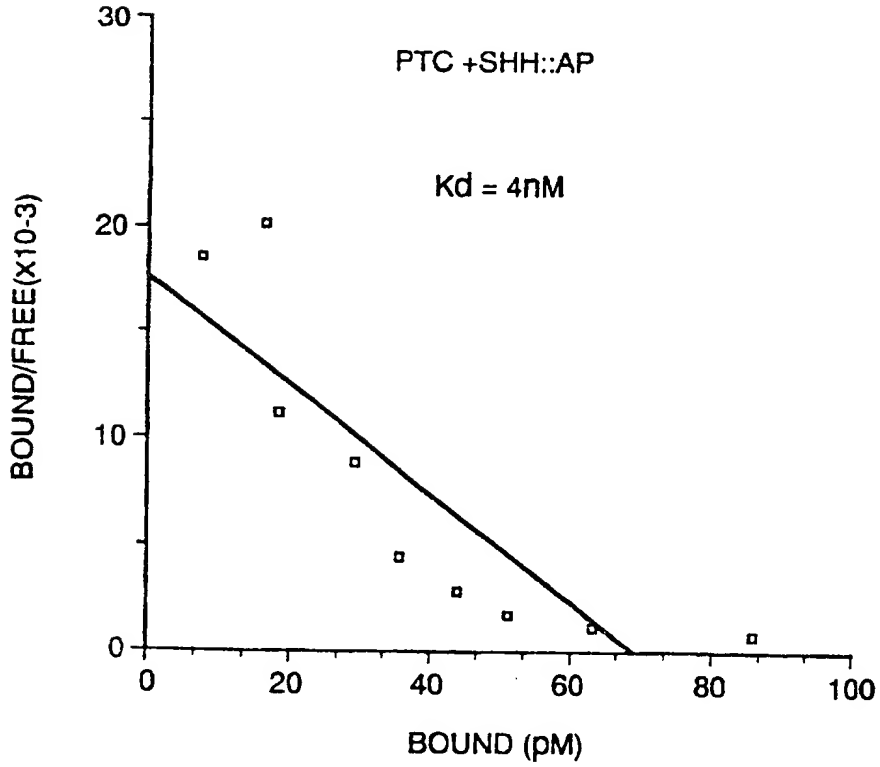

PTC +SHH::AP
Kd = 4nM
BOUND/FREE(X10-3)
BOUND (pM)
0
10
20
30
0
20
40
60
80
100

FIG. 3B

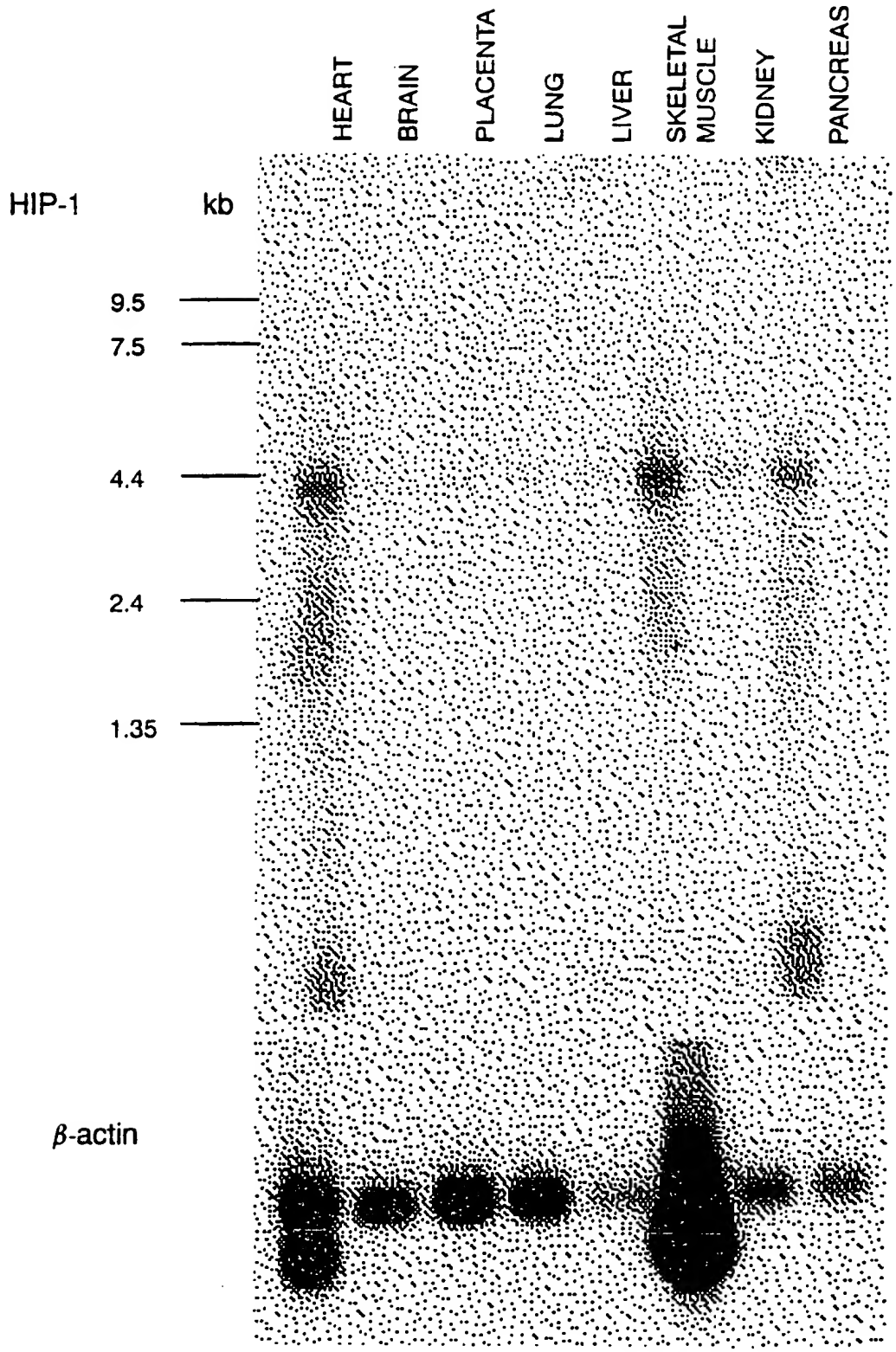

HEART
BRAIN
PLACENTA
LUNG
LIVER
SKELETAL MUSCLE
KIDNEY
PANCREAS
HIP-1
kb
9.5
7.5
4.4
2.4
1.35
β-actin

FIG. 4

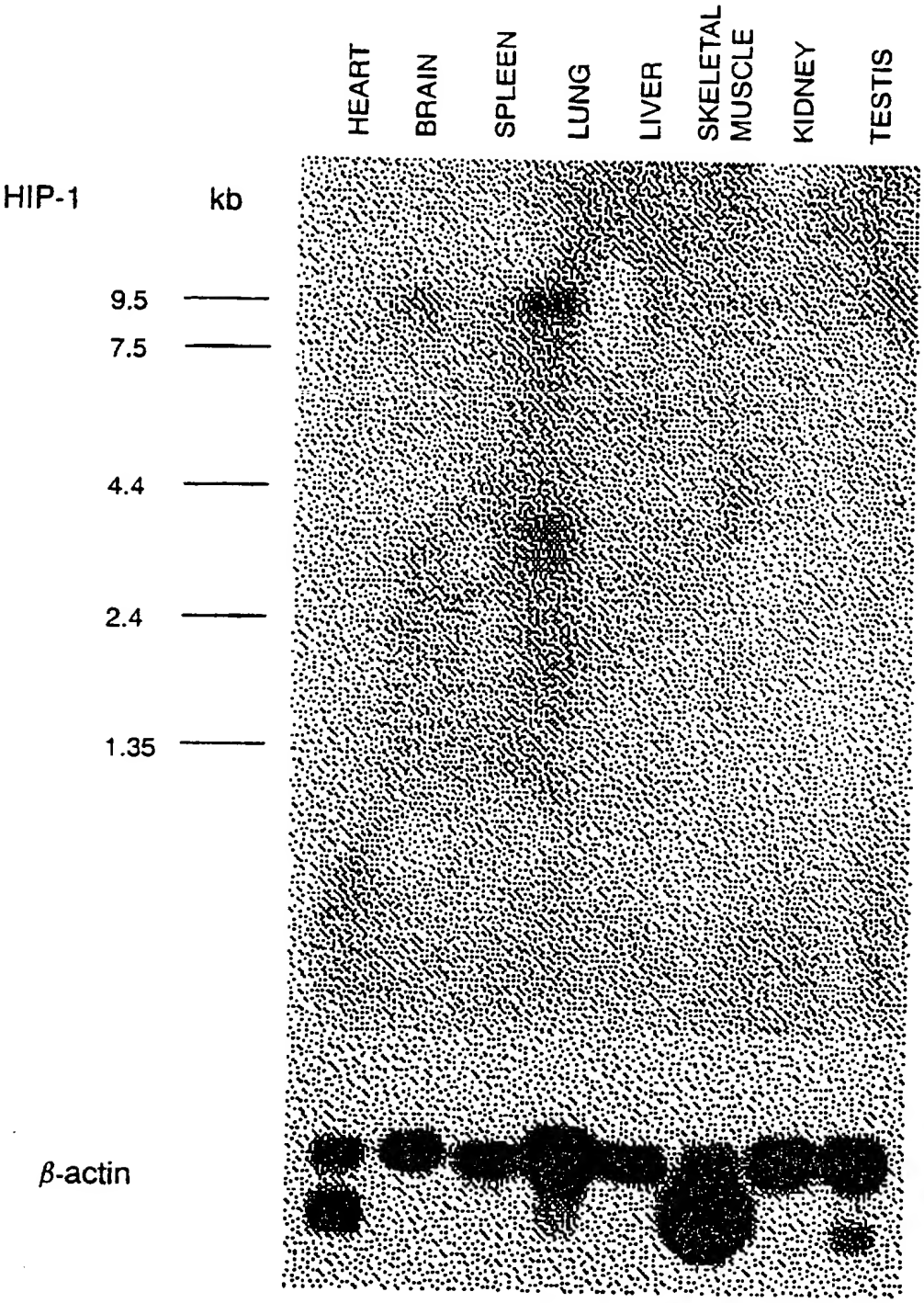
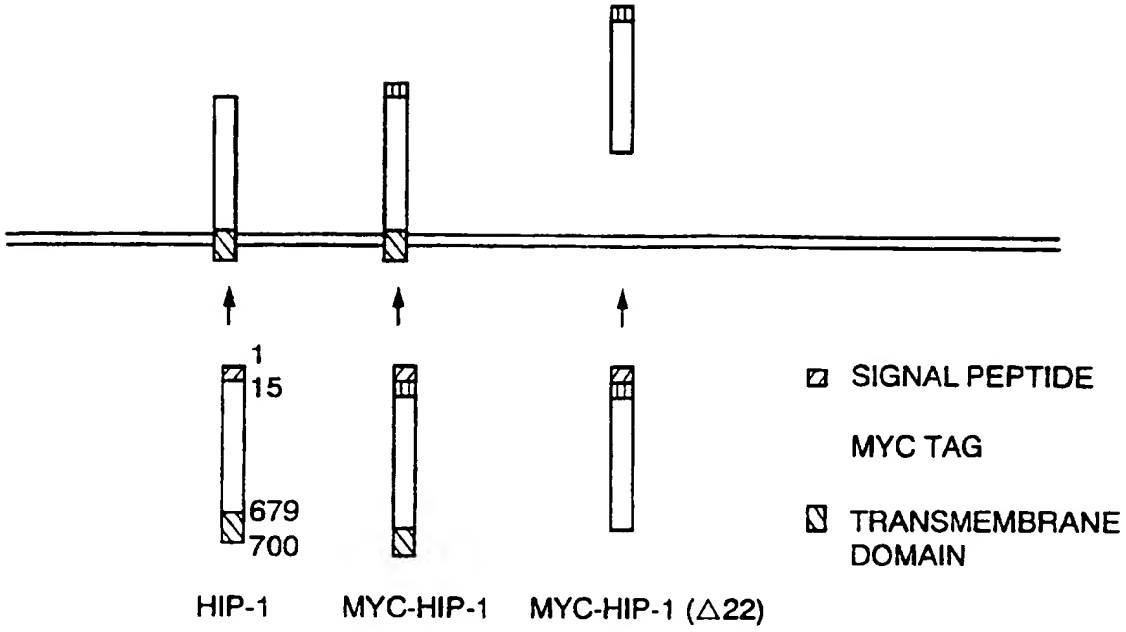

HEART
BRAIN
SPLEEN
LUNG
LIVER
SKELETAL MUSCLE
KIDNEY
TESTIS
HIP-1
kb
9.5
7.5
4.4
2.4
1.35
β-actin

FIG. 5


1
15
679
700
SIGNAL PEPTIDE
MYC TAG
TRANSMEMBRANE DOMAIN
HIP-1
MYC-HIP-1
MYC-HIP-1 (△22)

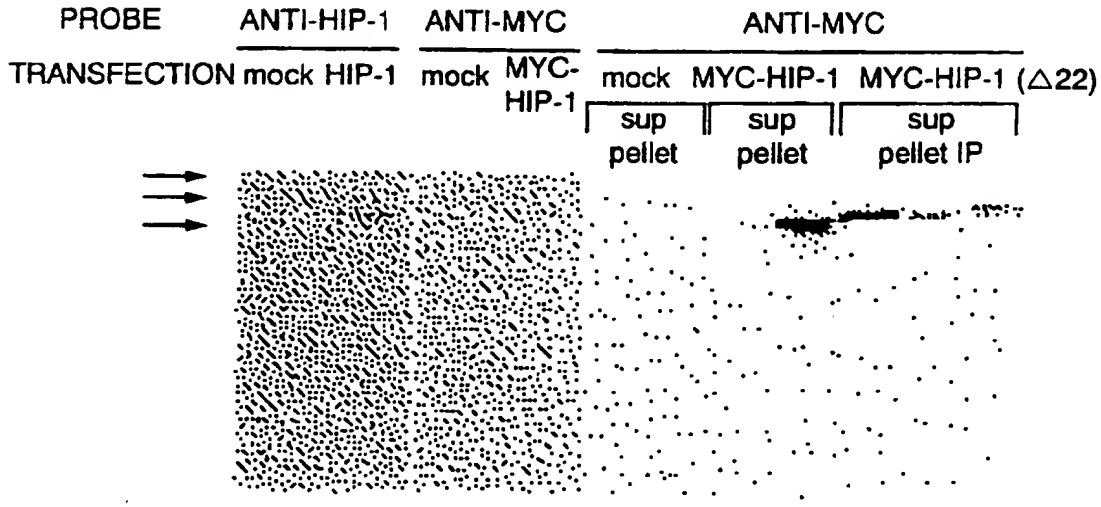

PROBE
ANTI-HIP-1
ANTI-MYC
ANTI-MYC
TRANSFECTION
mock
HIP-1
mock
MYC-HIP-1
mock
MYC-HIP-1
MYC-HIP-1 (△22)
sup
pellet
sup
pellet
sup
pellet
IP

FIG. 6

INTERNATIONAL SEARCH REPORT

Interna .al Application No
PCT/US 97/16741

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C07K14/46 C07K14/705 C07K16/18 C12N15/11
A01K67/027 C12Q1/68 G01N33/68 G01N33/566 A61K38/17
C12N15/62

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 95 18856 A (HARVARD COLLEGE ;IMP CANCER RES TECH (GB)) 13 July 1995 cited in the application see the whole document ---	
A	WO 96 17924 A (UNIV JOHNS HOPKINS MED ;UNIV WASHINGTON (US)) 13 June 1996 cited in the application see the whole document ---	
A	WO 96 16668 A (UNIV JOHNS HOPKINS MED ;UNIV WASHINGTON (US)) 6 June 1996 cited in the application see the whole document -----	

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

2

Date of the actual completion of the international search	Date of mailing of the international search report
12 February 1998	20.02.98

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Hix, R

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Int...ational application No.
PCT/US 97/16741

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

see FURTHER INFORMATION sheet PCT/ISA/210

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 97/16741

FURTHER INFORMATION C NTINUED FROM PCT/ISA/ 210

Remark : Although claims 48 to 52 are directed to a method of treatment of the human/animal body , the search has been carried out and based on the alleged effects of the compound/composition.